



2. VYDÁNÍ
AKTUALIZOVÁNO A
ROZŠÍŘENO

DATA DRIVEN CONSTRUCTION

NAVIGACE V DATOVÉM VĚKU VE
STAVEBNICTVÍ

S PŘÍPADY UŽITÍ AI A LLM

Artem Boiko



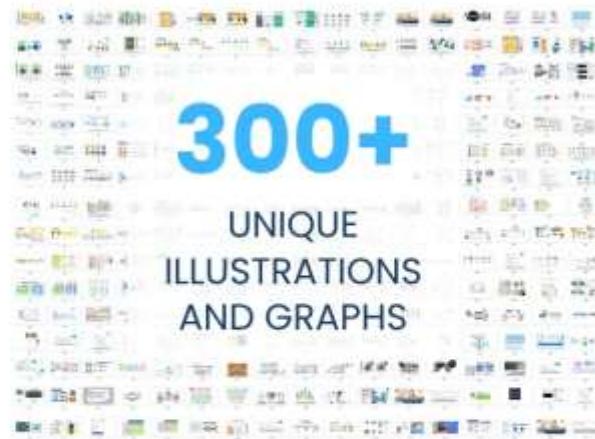
100+

KEY DATA-
RELATED
TOPICS



50+

PRACTICAL
PROBLEMS
SOLVED



300+

UNIQUE
ILLUSTRATIONS
AND GRAPHS

DATA-DRIVEN CONSTRUCTION

Navigace v době datové
ve stavebnictví

Druhé, přepracované a doplněné vydání

ARTEM BOIKO

“

"Boiko je James Carville IT - v jeho často citovaném "It's the economy, stupid" (Je to ekonomika, hloupá) je třeba pro tuto slavnou knihu vyměnit jen jedno slovo. "Jsou to data, hlupáku." (A aby se člověk v datovém vesmíru zorientoval, platí dodnes rčení starých Římanů pocházející z řečtiny: "Navigare necesse est". Autor naviguje své čtenáře všemi hlubinami i mělčinami datového oceánu jistou rukou a neochvějným kompasem, nemluvě o komplexním historickém přístupu a v neposlední řadě o velmi originální grafice a dobrém smyslu pro humor, který je patrný nejen na druhý pohled. Mezinárodní ohlasy na Boikovu knihu se pohybují od euforického souhlasu až po spíše žlučovitou skepsi, což druhému německému vydání knihy prospělo. Boiko je originální a nedogmatický myslitel dat. Předkládá čtenáři vzrušující postřehy a vždy odvážné, až provokativní teze, které inspirují k dalšímu přemýšlení. Vynikající lék na německou nemoc latentního konsenzualizmu. Mimochodem, výše uvedené latinské přísloví má doplněk: "vivere non est necesse". Na Bojkův přístup ke světu dat se nevztahuje - data žijí a jejich život je nutný, neřkuli zásadní."

- Dr. Burkhard Talebitari, nezávislý redaktor - mj. časopisu: BIM, který od roku 2013 každoročně vydává společnost Ernst & Sohn.

"Kniha Artyom Boika je milníkem pro demokratizaci digitalizace ve stavebnictví - a pro malé a střední podniky (MSP) představuje skutečnou změnu. Obzvláště průlomová: pomocí moderních open-source low-code a no-code nástrojů mohou firmy již nyní efektivně integrovat data do svých podnikových procesů a výhodně je analyzovat - bez hlubších znalostí programování. Tím se stává nákladné používání těžkopádných komerčních softwarových balíků zbytečným. Tato kniha je výzvou k akci! Je cenným průvodcem pro každého, kdo chce nejen porozumět digitální transformaci ve stavebnictví, ale chce ji také aktivně utvářet - pragmaticky, efektivně a perspektivně. Nyní je čas spolupracovat na sdílení těchto znalostí a trvale zvyšovat produktivitu stavebnictví."

- **Dr. Michael Max Buehler**, profesor stavebního managementu na HTWG Konstanz, spolumajitel společnosti GemeinWerk Ventures a nezávislý ředitel společnosti DevvStream.

"Kniha DataDrivenConstruction je jedním z prvních kroků za hranice běžného světa stavařů s jejich složitými systémy projektování a řízení, kdy, zdá se, složitost a přesycenost dat nedává šanci ani na radikální zjednodušení a zvýšení transparentnosti práce se stavebními daty. Artem ve své knize jednoduchým jazykem ukazuje, jaké možnosti před námi otevírají moderní technologie práce s daty, a doslova uvádí konkrétní kroky, které můžete okamžitě aplikovat ve své práci. Vyzývám všechny, kteří chtějí pochopit, kam se automatizační systémy ve stavebnictví posunou, aby si tuto knihu pozorně prostudovali a uvědomili si, že datová revoluce ve stavebnictví už nám klepe na dveře. Nyní je zajímavá jen pro geeky, ale za pár let, stejně jako BIM, budou takové přístupy a software všudypřítomné!"

- **Ihor Rogachew**, vedoucí kompetenčního centra IMT, BIM a digitální transformace ve společnosti RGD a zakladatel společnosti InfraBIM.Pro.

"Vřele doporučuji knihu DataDrivenConstruction, která se zabývá, jak říká název, přístupem k řízení informací na základě dat pro společnost AECO. V současné době ji používám k zahájení řady diskusí s různými skupinami. Zjistil jsem, že je to velmi přístupná příručka. Kromě důkladného přehledu historických souvislostí nástrojů v AECO, dat a představení několika klíčových technologií kniha obsahuje řadu velmi užitečných schémat, která nastínují rozsah datových zdrojů a artefaktů pro koncové uživatele s ukázkovými pracovními postupy. Zdá se mi, že právě tyto typy diagramů potřebujeme více při vývoji a monitorování informačních strategií a přispívají k BEP - definují celkový datový model podniku, na který lze překrýt hranice pro PIM a AIM."

- **Paul Ransley**, hlavní konzultant ve společnosti Acmena a inženýr systémové integrace ve společnosti Transport for London.

"Pokud jsou "data novou ropou", musíme se je naučit definovat, najít, vytěžit a zušlechtit, aby byla cenná. Kniha DataDrivenConstruction mi přišla velmi poučná a zasvěcená. Kniha poskytuje užitečné historické pozadí a srozumitelně vysvětluje práci s daty. Těm, kteří se zajímají o digitální transformaci, poskytne dobré pochopení dat - jak fungují, jak jsou strukturována a jak je lze využít."

- **Ralph Montague**, ředitel společnosti ArcDox, ředitel summitu koordinátorů BIM a předseda národního zrcadlového výboru BIM při irském národním úřadu pro normy.

"Jak bylo v knize zdůrazněno, informace jsou pro stavebnictví klíčovým přínosem a jejich dostupnost v přístupných formátech výrazně usnadňuje přesné rozhodování a zrychluje časový harmonogram projektu. Kniha nabízí neutrální a efektivní přístup k přístupu k tomuto zdroji a jeho využití při rozhodování."

Metodika představená v knize využívá moderní přístup, který kombinuje programování založené na umělé inteligenci s dostupnými nástroji s otevřeným zdrojovým kódem. Využitím sily umělé inteligence a využitím softwaru s otevřeným zdrojovým kódem si metodika klade za cíl zvýšit automatizaci, optimalizovat procesy a podpořit dostupnost a spolupráci v oboru. Jazyk knihy je srozumitelný a snadno pochopitelný."

- **Dr. Salih Ofluoğlu**, děkan Fakulty výtvarných umění a architektury na Antalya Bilim University a organizátor Euroasijského fóra BIM.

"Jediné, co mohu říct, je: WOW! Způsob, jakým jste začlenil historii, LLM, grafiku a celkovou snadnost pochopení vašich bodů, je opravdu pozoruhodný. Průběh knihy je úžasný. Tato kniha má totlik geniálních aspektů; skutečně mění pravidla hry. Je to skvělý zdroj informací a chválím vás za úsilí a nadšení, které jste do ní vložili. Gratuluji k vytvoření tak pozoruhodného díla. Mohl bych pokračovat, ale stačí říct, že jsem neuvěřitelně ohromen!"

- **Natalia Prinsloo**, vedoucí digitální praxe ve společnosti energylab_.

"Pro každého ve stavebnictví, od nováčků až po zkušené profesionály, je tato kniha pěvratná! Není to typické zaprášené čtení - je plná postřehů, strategií a špetky humoru, které vás zaujmou. Pokrývá vývoj využívání dat ve stavebnictví od starobylých metod zaznamenávání dat až po nejmodernější digitální technologie. Je to jako proletět se strojem času vývojem stavebních dat. Ať už jste architekt, inženýr, projektový manažer nebo datový analytik, tento komplexní průvodce způsobí revoluci ve vašem přístupu k projektům. Připravte se na optimalizaci procesů, zlepšení rozhodování a řízení projektů jako nikdy předtím!"

- **Pierpaolo Vergati**, přednášející na univerzitě Sapienza v Římě a vedoucí manažer stavebních projektů ve společnosti Fintecna.

"Knihu jsem přečetl jedním dechem za méně než 6 hodin. Kvalita zpracování knihy je vynikající, hustý lesklý papír, barevné provedení, příjemné písmo. Velké množství praktických příkladů, jak pracovat s LLM specifických pro stavebnictví, vám ušetří měsíce, ne-li roky samostudia. Příklady práce jsou velmi rozmanité, od jednoduchých až po složité, aniž byste museli kupovat složitý a drahý software. Kniha umožní majitelům jakéhokoli podniku ve stavebnictví nový pohled na jejich obchodní strategii, digitalizaci a perspektivy rozvoje. A menší firmám zvýšit efektivitu pomocí cenově dostupných a bezplatných nástrojů."

- **Mikhail Kosarev**, lektor a konzultant pro digitální transformaci ve stavebnictví ve společnosti TIM-ASG.

Kniha "DATA DRIVEN CONSTRUCTION" mění pravidla hry pro každého, koho zajímá, kam směřuje stavebnictví ve věku dat. Artem se nesnaží jen poškrábat povrch, ale proniká do hloubky současného vývoje, výzev a slibných příležitostí ve stavebnictví. Co tuto knihu odlišuje, je její přístupnost - Artem vysvětluje složité myšlenky pomocí srozumitelných analogií, které usnadňují pochopení obsahu. Kniha mi přišla neuvěřitelně poučná a zároveň poutavá. Celkově lze říci, že Artem vytvořil cenný zdroj informací, který nejen informuje, ale také inspiruje. Ať už jste zkušený profesionál, nebo nováček ve stavebnictví, tato kniha rozšíří váš rozhled a prohloubí vaše chápání toho, kam toto odvětví směřuje. Vřele doporučujeme!"

- **Moayad Saleh**, architekt a manažer implementace BIM ve společnosti TMM GROUP Gesamtplanungs GmbH.

"Měl bych říci, že kniha Data-Driven Construction je hodna toho, aby se vyučovala jako učebnice na vysokých školách, a je to kniha, která bude cenným přínosem pro rozvíjející se oblast BIM. Kniha Data-Driven Construction obsahuje technický slovníček, který velmi dobře vysvětluje pojmy. Témata, která je nesmírně obtížné vysvětlit, jsou zjednodušena a pochopena velmi krásným vizuálním jazykem. Myslím si, že to, co má být vysvětleno na vizuální stránce, by mělo být čtenáři vyjádřeno, i když stručně. Srozumitelnost některých vizuálů, jinými slovy čtení vizuálu vyžaduje samostatné informace. Chtěl bych také říci, že cenné práce Artema Boika rád představím na svých přednáškách a seminářích na univerzitách."

- **Dr. Ediz Yazicioglu**, majitel společnosti ArchCube a přednášející řízení stavebních projektů na katedře architektury Istanbulské technické univerzity a na univerzitě Medipol.

"Kniha Data Driven Construction názorně přibližuje základy práce se stavebními daty založené na informacích. Kniha, která se zabývá informačními toky a základními ekonomickými pojmy, se tak odlišuje od ostatních knih o BIM, protože nepředstavuje pouze pohled výrobce softwaru, ale snaží se zprostředkovat i základní pojmy. Kniha, kterou stojí za to si přečíst a vidět."

- **Jakob Hirn**, generální ředitel a spoluzakladatel společnosti Build Informed GmbH a iniciátor inovačního fóra "On Top With BIM".

"Data jsou nová ropa", jak se říká, a proto by jejich těžaři měli mít správné nástroje a myšlení, aby z tohoto zdroje 21. století vytěžili hodnotu. Stavebnictví se již příliš dlouho pohybuje na šíkmé ploše procesů založených na "3D informacích", kdy je dodávka projektu založena na informacích upečených někým jiným (např. již nakreslil koláčový nebo sloupový graf), zatímco podkladová "data" (např. nezpracovaná tabulka) jsou schopna přinést mnohem více, zejména proto, že fúze více dat a umělá inteligence přináší neomezený potenciál. Pokud dodáváte (nebo učíte/výzkumně pracujete) v oblasti stavebnictví, je tato kniha vaším nejlepším - a zatím jediným - zdrojem informací pro orientaci ve světě založeném na datech, v němž jsme se ocitli."

- **Dr. Zulfikar Adamu**, docent strategického IT ve stavebnictví na LSBU, Velká Británie.

"Data-Driven Construction" od Artema Boika je působivé dílo, které nabízí solidní základ pro stavebnictví v době neustále rostoucích technologií a informačních možností. Boiko dokáže srozumitelně představit složitá téma a zároveň přináší vizizánské myšlenky. Kniha je promyšleným kompendiem, které nejen upozorňuje na současný vývoj, ale také poskytuje výhled na budoucí inovace. Vřele ji doporučujeme všem, kteří chtějí proniknout do problematiky plánování a realizace staveb založených na datech."

- **Markus Eiberger**, přednášející na Vysoké škole aplikovaných věd ve Stuttgartu, vedoucí projektový manažer a zástupce vedoucího pobočky ve společnosti Konstruktionsgruppe Bauen, člen představenstva sdružení BIM Cluster Baden-Württemberg.



Druhé vydání, březen 2025.
© 2025 | Artem Boiko | Karlsruhe

ISBN 978-3-912002-07-2



Artem Boiko Autorská práva

boikoartem@gmail.com
info@datadrivenconstruction.io

Žádná část této knihy nesmí být reprodukována nebo přenášena v jakékoli formě nebo jakýmkoli prostředky, elektronickými nebo mechanickými, včetně fotokopírování, nahrávání nebo jakéhokoli systému pro ukládání a vyhledávání informací, bez písemného souhlasu autora - s výjimkou nekomerčního šíření v nezměněné podobě. Knihu je šířena zdarma a může být volně sdílena s dalšími uživateli pro osobní, vzdělávací nebo výzkumné účely za předpokladu, že bude zachováno autorství a odkazy na originál. Autor si ponechává veškerá nevlastnická práva k textu a neposkytuje žádné výslovné ani implicitní záruky. Společnosti, produkty a názvy uvedené v knize mohou být fiktivní nebo použité jako příklady. Autor nenese odpovědnost za jakékoli důsledky použití uvedených informací. Informace obsažené v knize jsou poskytovány "tak, jak jsou", bez záruký uplnosti nebo relevance. Autor neodpovídá za náhodné nebo následné škody vzniklé v důsledku použití informací, kódů nebo programů obsažených v této knize. Ukázky kódů uvedené v této knize slouží pouze pro vzdělávací účely. Čtenář je používají na vlastní nebezpečí. Autor doporučuje, aby všechna softwarová řešení byla před použitím v produkčním prostředí otestována. Všechny ochranné známky a názvy produktů uvedené v textu jsou ochranné známky, registrované ochranné známky nebo značky služeb příslušných společností a jsou majetkem příslušných vlastníků. Použití těchto názvů v knize neznamená žádný vztah s jejich vlastníky ani jejich podporu. Zmínky o produktech nebo službách třetích stran nepředstavují doporučení ani neznamenají jejich podporu. Názvy společností a produktů použité v příkladech mohou být ochrannými známkami jejich vlastníků. Odkazy na webové stránky třetích stran jsou uvedeny pro usnadnění a neznamenají, že autor schvaluje informace uvedené na těchto stránkách. Všechny uvedené statistiky, citace a studie byly aktuální v době psaní článku. Údaje se mohou v průběhu času měnit.

Tato kniha je šířena pod licencí Creative Commons Uveďte autora-Neužívejte komerčně-Nezpracovávejte 4.0 International (CC BY-NC-ND 4.0). Můžete ji kopírovat a šířit pro nekomerční účely za předpokladu, že si zachováte autorství a nebudete provádět žádné změny.



CC BY-NC-ND 4.0

© 2024 Artyom Boiko. První vydání.
© 2025 Artyom Boiko. Druhé, přepracované a doplněné vydání.
Všechna práva vyhrazena.

PŘEDMLUVA K DRUHÉMU VYDÁNÍ

Tato kniha je výsledkem živého dialogu s odbornou veřejností. Vychází z četných odborných diskusí o správě dat ve stavebnictví, které probíhaly na různých odborných platformách a sociálních sítích. Tyto diskuse se staly základem pro články, publikace a obrazové materiály, které vyvolaly široký ohlas v odborné komunitě. Autorův obsah každoročně přitahuje miliony zhlédnutí na různých platformách a v různých jazycích a sdružuje odborníky v oblasti digitalizace stavebnictví.

Do roka od vydání prvního vydání si knihu objednali odborníci z více než 50 zemí, od Brazílie a Peru po Mauricius a Japonsko. Druhé vydání knihy, které právě držíte v rukou, bylo přepracováno a rozšířeno na základě zpětné vazby odborníků, kritiky prvního vydání a diskusí v odborných kruzích. Díky zpětné vazbě bylo druhé vydání výrazně rozšířeno: přibyly nové kapitoly o technologiích CAD (BIM) a o vytváření efektivních ETL -procesů. Výrazně se také zvýšil počet praktických příkladů a případových studií. Zvláště cenná je zpětná vazba od vedoucích pracovníků stavebního průmyslu, poradenských společností a významných IT-společností, kteří se na autora obraceli s dotazy na digitalizaci a interoperabilitu jak před vydáním první verze knihy, tak po něm. Mnozí z nich již přistupy popsané v knize použili nebo tak plánují učinit v blízké budoucnosti

Držíte v rukou knihu, která vznikla na základě diskuse a aktivní výměny názorů. Pokrok se rodí v dialogu, ve střetu názorů a otevřenosti novým přístupům. Děkujeme vám, že jste součástí tohoto dialogu. Vaše konstruktivní kritika je základem pro budoucí zlepšení. Pokud v textu najdete chyby nebo se chcete podělit o nápady a návrhy, je vítána jakákoli zpětná vazba. Kontaktní údaje jsou uvedeny na konci knihy

PROČ JE KNIHA ZDARMA?

Tato kniha byla koncipována jako otevřený vzdělávací zdroj zaměřený na šíření moderních přístupů ke správě dat ve stavebnictví. První verze knihy posloužila jako základ pro sběr připomínek a podnětů od odborné veřejnosti, což umožnilo zlepšit strukturu a obsah materiálu. Všechny připomínky, návrhy a nápady byly pečlivě analyzovány a zapracovány do této revidované verze. Cílem knihy je pomocí stavebním odborníkům pochopit význam práce s daty: systematicky, vědomě a s ohledem na dlouhodobou hodnotu informací. Autor shromáždil příklady, ilustrace a praktické postřehy z více než desetileté praxe v oblasti digitalizace stavebnictví. Většina tohoto materiálu vznikla na základě skutečných projektů, diskusí s inženýry a developery, účasti na mezinárodních iniciativách a vzdělávacích seminářích. Kniha je pokusem o strukturování nashromážděných zkušeností a jejich sdílení přístupnou formou. Pokud chcete podpořit další šíření myšlenek knihy a získat pohodlný formát pro čtení, práci s příklady a obrazovými materiály - můžete si zakoupit [tištěnou verzi](#).

PRÁVA NA UŽÍVÁNÍ

Všechny materiály, ilustrace a fragmenty této knihy lze reprodukovat, citovat nebo používat v jakémkoli formátu a na jakémkoli médiu pod podmírkou, že bude uveden zdroj: autorství Artem Boiko a název knihy "Data-Driven Construction". Děkujeme za úctu k práci a šíření znalostí.

Tuto knihu s upřímnou vděčností věnuji své rodině, která mi od
útlého věku vštěpovala hlubokou lásku ke stavitelství, svému
rodnému hornickému městu za lekce odolnosti a své manželce,
která je mi stálou inspirací.

PRO KOHO JE TATO KNIHA URČENA

Kniha je psána přístupným jazykem a je určena širokému okruhu čtenářů ve stavebnictví - od studentů a začátečníků, kteří chtějí pochopit základy moderních stavebních procesů, až po profesionály, kteří potřebují aktuální metodiku pro správu dat ve stavebnictví. Ať už jste architekt, inženýr, mistr, stavební manažer nebo datový analytik, tato komplexní příručka s mnoha unikátními ilustracemi a grafikou nabízí cenné poznatky o tom, jak využívat data v podnikání k optimalizaci a automatizaci procesů, zlepšení rozhodování a řízení stavebních projektů na různých úrovních pomocí moderních nástrojů.

Kniha je komplexním průvodcem, který kombinuje teoretické základy a praktická doporučení pro integraci technik správy dat do stavebních procesů. Kniha se zaměřuje na strategické využití informací pro optimalizaci provozu, automatizaci procesů, zlepšení rozhodování a efektivní řízení projektů pomocí moderních digitálních nástrojů.

Tato kniha se zabývá teoretickými a praktickými aspekty práce s informacemi ve stavebnictví. Na podrobných příkladech se zabývá metodikou parametrizace úloh, shromažďováním požadavků, zpracováním nestrukturovaných a víceformátových dat a jejich transformací do efektivních řešení pro stavební firmy.

Čtenář postupně projde cestu od tvorby požadavků a vývoje základních datových modelů ke složitějším procesům integrace heterogenních informačních zdrojů, tvorbě ETL procesů, konstrukci informačních Pipelines a modelů strojového učení. Postupný přístup umožňuje názorně ukázat mechanismy organizace a automatizace podnikových procesů a systémů pro podporu rozhodování ve stavebnictví. Každou část knihy uzavírá praktická kapitola obsahující návody krok za krokem, které umožňují okamžitě aplikovat získané znalosti v reálných projektech.

SHRNUTÍ ČÁSTÍ KNIHY

Tato kniha je strukturována kolem konceptu transformace dat v hodnotovém řetězci: od sběru dat a zajištění jejich kvality až po analytické zpracování a získání cenných praktických řešení pomocí moderních nástrojů a metodik.

Část 1: Digitální vývoj ve stavebnictví - sleduje historickou proměnu správy dat od hliněných tabulek k moderním digitálním systémům, analyzuje vznik modulárních systémů a rostoucí význam digitalizace informací v kontextu průmyslových revolucí.

Část 2: Informační výzvy pro stavebnictví - zkoumá problémy roztríštěnosti dat, "informačních sil", dopad přístupu HiPPO rozhodování a omezení proprietárních formátů a navrhuje zvážit přechod na ekosystémy umělé inteligence a LLM.

Část 3: Systematizace dat ve stavebnictví - vytváří typologii stavebních dat, popisuje metody jejich organizace, integrace s podnikovými systémy a zabývá se vytvářením kompetenčních center pro standardizaci informačních procesů.

Část 4: Zajištění kvality dat - odhaluje metodiky pro přeměnu různorodých informací na kvalitní strukturovaná data, včetně extrakce dat z různých zdrojů, validace a modelování pomocí LLM.

Část 5: Kalkulace nákladů a času - zabývá se digitalizací kalkulací nákladů a plánování, automatizací získávání objemů z modelů CAD (BIM), technologiemi modelování 4D-8D a výpočty ESG stavebních projektů.

Část 6: CAD a BIM - kriticky analyzuje vývoj technologií pro navrhování, otázky interoperability systémů, trendy směrem k otevřeným datovým formátům a vyhlídky na využití umělé inteligence v navrhování.

Část 7: Analýza dat a automatizace - věnuje se principům vizualizace informací, klíčovým ukazatelům výkonnosti, procesům ETL, nástrojům pro orchestraci pracovních postupů a použití jazykových modelů k automatizaci rutinních úloh.

Část 8: Ukládání a správa dat - zkoumá formáty ukládání dat, koncepty datových skladů a datových jezer, zásady správy dat a nové přístupy včetně vektorových databází a metodik DataOps a VectorOps.

Část 9: Big Data a strojové učení - zaměřuje se na přechod k objektivní analýze založené na historických datech, internet věcí na stavbách a použití algoritmů strojového učení k předvídání nákladů a časového harmonogramu projektů.

Část 10: Stavebnictví ve věku digitálních dat - představuje pohled na budoucnost stavebnictví, analyzuje posun od kauzální analýzy k práci s korelacemi, koncept "uberizace" stavebnictví a strategie digitální transformace.

What is meant by **data-driven construction** ?



ÚVOD

Jak dlouho může vaše společnost zůstat konkurenceschopná ve světě, kde se technologie rychle vyvíjejí a každý aspekt podnikání, od časového plánování a kalkulace nákladů až po analýzu rizik, je automatizován modely strojového učení?

Stavebnictví, které existuje stejně dlouho jako lidstvo samo, je na prahu revoluční změny, která slibuje, že zcela změní nás pohled na tradiční stavebnictví. Již v jiných odvětvích ekonomiky digitalizace nejen mění pravidla, ale nemilosrdně vytlačuje z trhu společnosti, které se nedokázaly přizpůsobit novému prostředí zpracování dat a nejsou schopny zvýšit rychlosť rozhodování (obr. 1).



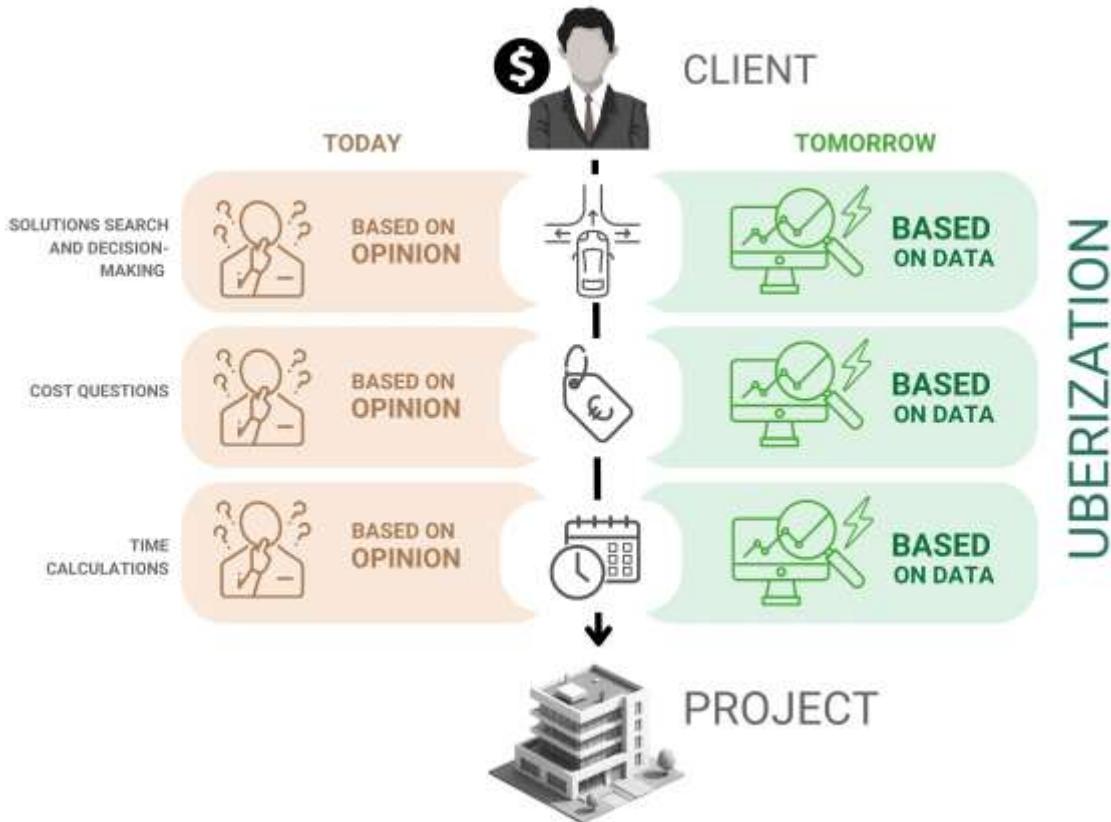
Obr. 1 Rychlosť rozhodovania ve stavebnictví závisí na lidskom faktore častejnež v jiných odvětvích.

Bankovnictví, maloobchod, logistika a zemědělství rychle směřují k plné digitalizaci, kde již nemají místo nepřesnosti a subjektivní názory. Moderní algoritmy jsou schopny analyzovat obrovské množství dat a poskytovat zákazníkům přesné předpovědi - ať už jde o pravděpodobnost splacení úvěru, optimální trasy dodávek nebo předpověď rizik.

Stavebnictví je jedním z posledních odvětví, které nevyhnutelně přechází od řešení založených na názorech vysoce placených odborníků k řešením založeným na datech. Tento přechod je poháněn nejen novými technologickými možnostmi, ale také zvýšenými požadavky trhu a zákazníků na transparentnost, přesnost a rychlosť.

Robotizace, automatizace procesů, otevřená data a na nich založené prognózy - to vše už není jen možnost, ale nevyhnutelnost. Většině firem ve stavebnictví, které byly ještě nedávno vůči klientovi zodpovědné za kalkulaci objemu, nákladů, času projektů a kontrolu kvality, nyní hrozí, že se změní v pouhé vykonavatele zakázek, kteří nebudou činit klíčová rozhodnutí (obr. 2).

Díky pokroku ve výpočetním výkonu, algoritmům strojového učení a demokratizovanému přístupu k datům je nyní možné automaticky kombinovat data z různých zdrojů, což umožňuje hlubší analýzu procesů, předvídání rizik a optimalizaci nákladů v diskusních fázích stavebního projektu. Tyto technologie mají potenciál radikálně zvýšit efektivitu a snížit náklady v celém odvětví.



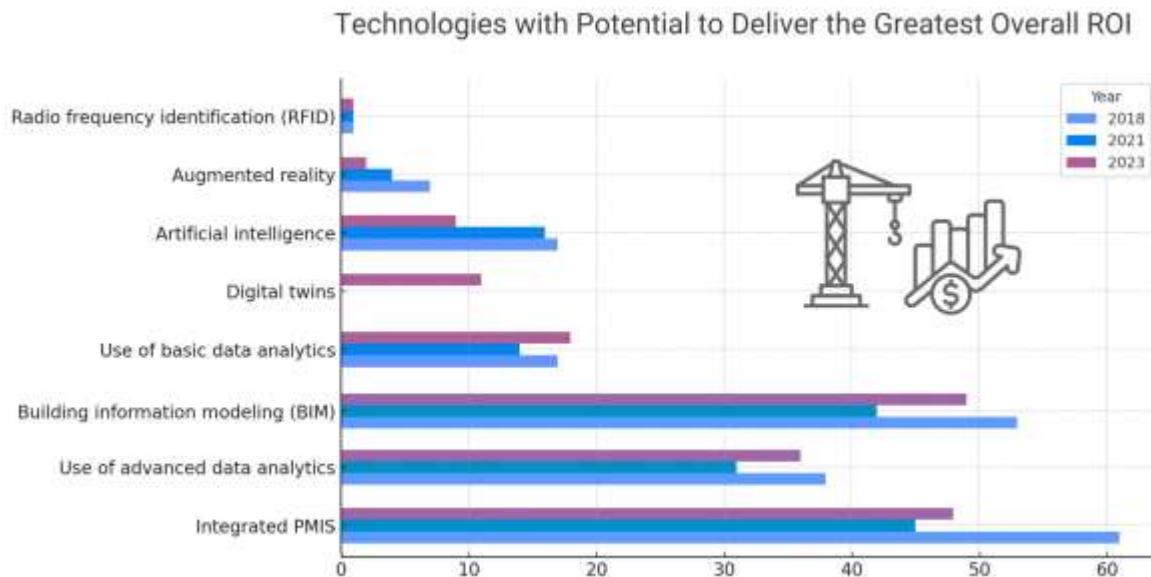
Obr. 2 Zákazník nemá zájem o nadměrný lidský faktor na cestě k realizaci svého projektu.

Navzdory všem výhodám nových nástrojů a koncepcí stavebnictví v zavádění nových technologií značně zaostává za ostatními hospodářskými odvětvími.

Podle zprávy IT Metrics Key Data 2017 je stavebnictví na posledním místě ve výdajích na IT mezi 19 dalšími odvětvími [1].

Rychlý nárůst objemu dat a složitosti procesů se stává pro vedení společnosti bolestí hlavy a hlavním problémem při využívání nových technologií je, že data i přes jejich množství zůstávají roztríštěná, nestrukturovaná a často nekompatibilní mezi různými systémy a softwarovými produkty. Proto se dnes mnoho společností ve stavebnictví zabývá především otázkami kvality dat, které lze vyřešit pouze zavedením efektivních, automatizovaných systémů pro správu a analýzu.

Podle průzkumu, který provedla společnost KPMG® mezi stavebními manažery v roce 2023 [2], mají největší potenciál zlepšit návratnost investic do projektů informační systémy pro řízení projektů (PMIS), pokročilá a základní datová analýza a informační modelování budov (BIM) (obr. 3).



Obr. 3 Průzkum mezi manažery stavebních firem: Které technologie zajistí nejvyšší návratnost investic (ROI) do investičních projektů? (na základě materiálů [2]).

Řešením problémů spojených s integrací dat do podnikových procesů je zajištění vysoké kvality informací, používání vhodných datových formátů a používání efektivních metod pro vytváření, ukládání, analýzu a zpracování dat.

Uvědomení si hodnoty dat nutí různá průmyslová odvětví opustit oddělené aplikace a složité byrokratické struktury řízení. Místo toho se pozornost přesouvá na vytváření nových přístupů k informační architektuře, které mění společnosti v moderní podniky založené na datech. Dříve či později tento krok učiní i samotné stavebnictví, které se od postupné digitální evoluce posune ke skutečné digitální revoluci, jež se dotkne všech společností.

Přechod na podnikové procesy založené na datech nebude snadný. Mnoho společností bude čelit problémům, protože vedoucí pracovníci nevždy chápou, jak využít chaotické soubory dat ke zvýšení efektivity a růstu podniku.

Tato kniha se zabývá světem dat, kde se informace stávají klíčovým strategickým zdrojem, který určuje efektivitu a udržitelnost podnikových procesů. S rychlým růstem objemu informací čelí společnosti novým výzvám. Digitální transformace již není jen módním slovem - stává se nutností.



Obr. 4 Základem konstrukce jsou data a procesy.

Porozumět transformaci znamená umět vysvětlit složité věci jednoduchými slovy. Proto je kniha na psána přístupným jazykem a je doplněna autorovými ilustracemi, vytvořenými speciálně pro jasné vysvětlení klíčových pojmů. Tato schémata, grafy a vizualizace jsou navrženy tak, aby odstranily překážky ve vnímání a učinily látku srozumitelnou i těm, kteří dříve považovali taková téma za příliš složitá. Všechny ilustrace, diagramy a grafy v této knize jsou vytvořeny autorem a jsou navrženy speciálně pro vizualizaci klíčových pojmů popsaných v textu.

Jeden obrázek vydá za tisíc slov [3].

- Fred R. Barnard, anglický ilustrátor, 1927.

Abychom propojili teorii s praxí, budeme používat nástroje umělé inteligence (zejména jazykové modely), které vám umožní vyvíjet řešení bez nutnosti hlubokých znalostí programování. Pokud jste orientováni na praktický materiál a zajímá vás spíše praktická práce s daty, můžete první úvodní část přeskočit a přejít rovnou k druhé části knihy, kde začíná popis konkrétních příkladů a případů.

Nevkládejte však přílišná očekávání do nástrojů umělé inteligence, strojového učení a LLM (Large Language Models) obecně. Bez kvalitních vstupních dat a hlubokého porozumění dané problematice nemohou ani ty nejpokročilejší algoritmy poskytnout spolehlivé a smysluplné výsledky.

Generální ředitel společnosti Microsoft Satya Nadella varuje před rizikem bubliny v oblasti umělé inteligence na počátku roku 2025 [4] a přirovnává současný humbuk k bublině dot-com. Zdůrazňuje, že tvrzení o dosažení milníků AGI (Artificial General Intelligence) bez rádného zdůvodnění jsou "nesmyslnou manipulací s metrikami". Nadella se domnívá, že skutečný úspěch umělé inteligence by se měl měřit spíše podle jejího přínosu k růstu globálního HDP než podle přehnaného zdůrazňování módních slov.

Za všemi hesly o nových technologiích a koncepcích se skrývá složitá a pečlivá práce na zajištění

kvality dat, parametrizace podnikových procesů a přizpůsobení nástrojů reálným úlohám.

Přístup založený na datech není produkt, který si můžete jen tak stáhnout nebo koupit. Je to strategie, kterou je třeba vybudovat. Začíná novým pohledem na stávající procesy a problémy a poté vyžaduje disciplinovaný pohyb zvoleným směrem.

Přední vývojáři softwaru a dodavatelé aplikací nebudou motorem změn ve stavebnictví, pro mnohé z nich je přístup založený na datech hrozbou pro jejich zavedený obchodní model.

Jiná odvětví [na rozdíl od stavebnictví], například automobilový průmysl, již prošla radikální a převratnou změnou a jejich digitální transformace je v plném proudu. Stavební společnosti musí jednat rychle a rozhodně: hbité společnosti získají obrovské výhody, zatímco pro ty, které váhají, budou rizika vážná. Vzpomeňte si na převrat, který v tomto odvětví způsobila digitální fotografie [5].

- Zpráva Světového ekonomického fóra Shaping the Future of Construction, 2016.

Společnosti, které včas rozpoznají příležitosti a výhody nového přístupu, získají trvalou konkurenční výhodu a budou se mocí rozvíjet a růst bez závislosti na řešeních od velkých dodavatelů.

To je vaše šance nejen přečkat nadcházející bouři digitalizace informací, ale i převzít nad ní kontrolu. V této knize najdete nejen analýzu současného stavu oboru, ale také konkrétní doporučení, jak přehodnotit a restrukturalizovat své procesy a podnikání, abyste se stali lídrem v nové éře stavebnictví a zvýšili své profesní zkušenosti.

Digitální budoucnost stavebnictví není jen o používání nových technologií a programů, ale o zásadním přehodnocení nakládání s daty a obchodních modelů.

Je vaše společnost připravena na tuto strategickou změnu?

OBSAH

ÚVOD.....	1
OBSAH	I
II ČÁST OD HLINĚNÝCH TABULEK K DIGITÁLNÍ REVOLUCI: JAK SE VYVÍJELY INFORMACE VE STAVEBNICTVÍ.....	2
KAPITOLA 2.1. VÝVOJ VYUŽÍVÁNÍ DAT VE STAVEBNICTVÍ	3
Zrod datové éry ve stavebnictví.....	3
Od hlíny a papyru k digitální technologii	4
Proces jako nástroj pro zkušenosti založené na datech	5
Digitalizace informací o stavebním procesu	7
KAPITOLA 2.2. TECHNOLOGIE A SYSTÉMY ŘÍZENÍ V MODERNÍM STAVEBNICTVÍ.....	11
Digitální revoluce a vznik modulárních MRP/ERP -systémů	11
Systémy pro správu dat: od dolování dat k obchodním výzvám	13
Firemní mycelium: jak se data propojují s podnikovými procesy	16
KAPITOLA 2.3. DIGITÁLNÍ REVOLUCE A EXPLOZE DAT.....	20
Počátek datového boomu jako evoluční vlna	20
Množství dat generovaných v moderní společnosti	22
Náklady na ukládání dat: ekonomické hledisko.....	23
Hranice akumulace dat: od hmoty k významu.....	25
Další kroky: od teorie dat k praktickým změnám.....	27
III ČÁST JAK SE STAVEBNICTVÍ UTÁPÍ V DATOVÉM CHAOSU.....	28
KAPITOLA 3.1. ROZTŘÍSTĚNOST DAT A DATOVÁ SILA.....	29
Čím více nástrojů, tím efektivnější podnikání?	29
Datová sila a jejich dopad na výkonnost společnosti.....	31
Duplicita a nedostatečná kvalita dat jako důsledek nejednotnosti.....	34
HiPPO aneb nebezpečí názorů při rozhodování.....	36
Neustálý nárůst složitosti a dynamiky podnikových procesů	39
Čtvrtá průmyslová revoluce (Průmysl 4.0) a pátá průmyslová revoluce (Průmysl 5.0) ve stavebnictví.	42
KAPITOLA 3.2. PŘEMĚNA CHAOSU V ŘÁD A SNÍŽENÍ SLOŽITOSTI	45
Zbytečný kód a uzavřené systémy jako překážka zvýšení produktivity	45
Od silosů k jednotnému datovému skladu.....	47

Integrované úložné systémy umožňují přechod na agenty s umělou inteligencí.....	48
Od sběru dat k rozhodování: cesta k automatizaci	51
Další kroky: přeměna chaosu na zvládnutelný systém	53
IV ČÁST DATOVÝ RÁMEC V PODNIKOVÝCH PROCESECH VE STAVEBNICTVÍ 55	
KAPITOLA 4.1. DATOVÉ TYPY VE STAVEBNICTVÍ.....	56
Nejdůležitější typy dat ve stavebnictví.....	56
Strukturovaná data	60
Relační databáze RDBMS a dotazovací jazyk SQL	61
SQL - dotazy v databázích a nové trendy	64
Nestrukturovaná data	66
Textová data: mezi nestrukturovaným chaosem a strukturovaným chaosem y	67
Polostrukturovaná a volně strukturovaná data	68
Geometrická data a jejich použití	69
Data CAD: od návrhu po ukládání dat.....	72
Vznik konceptu BIM (BOM) a využití CAD v procesech.....	75
KAPITOLA 4.2. HARMONIZACE A STRUKTUROVÁNÍ DAT.....	81
Naplňování systémů daty ve stavebnictví.....	81
Transformace dat: kritický základ moderní podnikové analýzy	83
Datové modely: vztahy v datech a vztahy mezi prvky	87
Proprietární formáty a jejich dopad na digitální procesy	91
Otevřené formáty mění přístup k digitalizaci.....	95
Změna paradigmatu: Open Source jako konec éry dominance dodavatelů softwaru	96
Strukturovaná otevřená data: základ digitální transformace	99
KAPITOLA 4.3. LLM A JEJICH ÚLOHA PŘI ZPRACOVÁNÍ DAT A OBCHODNÍCH PROCESECH.....	102
Chatovací místo LLM: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok k automatizaci datových procesů.....	102
Velké jazykové modely LLM: jak to funguje	103
Využití místních LLM pro citlivé firemní údaje	106
Plná kontrola nad AI ve firmě a způsob nasazení vlastního LLM	108
RAG: Inteligentní LLM -asistenti s přístupem k firemním datům.....	110
KAPITOLA 4.4. IDE S PODPOROU LLM A BUDOUCÍ PROGRAMOVÉ ZMĚNY	112
Výběr IDE: od experimentů LLM po podniková řešení.....	112
IDE s podporou LLM a budoucí programové změny	114
Python Pandas: nepostradatelný nástroj pro práci s daty	115

DataFrame: univerzální formát tabulkových dat	119
Další kroky: budování udržitelného datového rámce	122
V ČÁST KVALITA DAT: ORGANIZACE, STRUKTUROVÁNÍ, MODELOVÁNÍ.....	124
KAPITOLA 5.1. PŘEVOD DAT DO STRUKTUROVANÉ PODOBY.....	125
Naučit se převádět dokumenty, PDF, obrázky a texty do strukturovaných formátů.....	125
Příklad převodu dokumentu PDF na tabulku.....	126
Převod obrázku JPEG, PNG do strukturované podoby	130
Převod textových dat do strukturované podoby	133
Převod dat CAD (BIM) do strukturované podoby	136
Dodavatelé řešení CAD směřují ke strukturovaným datům	141
KAPITOLA 5.2. KLASIFIKACE A INTEGRACE: SPOLEČNÝ JAZYK PRO STAVEBNÍ DATA	144
Rychlosť rozhodování závisí na kvalitě dat.....	144
Standardizace a integrace dat.....	145
Digitální interoperabilita začíná u požadavků.....	148
Společný jazyk konstrukce: úloha klasifikátorů v digitální transformaci	150
Masterformat, OmniClass, Uniclass a CoClass: vývoj klasifikačních systémů.....	153
KAPITOLA 5.3. DATOVÉ MODELOVÁNÍ A CENTRUM EXCELENCE	158
Modelování dat: konceptuální, logický a fyzický model	158
Praktické modelování dat v kontextu stavebnictví	161
Vytvoření databáze pomocí LLM	164
Centrum excelence (CoE) pro datové modelování	166
KAPITOLA 5.4. SYSTEMATIZACE POŽADAVKŮ A OVĚŘOVÁNÍ INFORMACÍ.....	169
Shromažďování a analýza požadavků: transformace komunikace do strukturovaných dat.	169
Procesní schémata a účinnost koncepčních rámců	173
Strukturované požadavky a regulární výrazy RegEx	175
Sběr dat pro proces ověřování.....	180
Ověření údajů na a výsledky ověření	182
Vizualizace výsledků ověřování.....	187
Srovnání kontrol kvality dat s potřebami lidského života	189
Další kroky: proměna dat v přesné výpočty a plány	191
VI ČÁST KALKULACE NÁKLADŮ A ČASU: ZAČLENĚNÍ DAT DO STAVEBNÍCH PROCESŮ.....	193
KAPITOLA 6.1. VÝPOČTY A ODHADY NÁKLADŮ NA STAVEBNÍ PROJEKTY	194
Základy stavebnictví: odhad množství, nákladů a času	194
Metody výpočtu odhadovaných nákladů projektů	195

Metoda založená na zdrojích kalkulace nákladů a odhadu ve stavebnictví.....	196
Databáze stavebních zdrojů: katalog stavebních materiálů a prací.....	196
Sestavení kalkulací a vyčíslení nákladů na práce na základě zdrojové základny.	198
Konečné vyčíslení nákladů projektu: od odhadu k rozpočtu	203
KAPITOLA 6.2. VÝPOČET MNOŽSTVÍ A AUTOMATICKÉ VYTVAŘENÍ ODHADŮ A HARMONOGRAMŮ ..	207
Přechod z 3D na 4D a 5D: využití objemových a kvantitativních parametrů	207
5D atributy a získávání objemů atributů z CAD.....	207
QTO Quantity Take-Off: seskupení dat projektu podle atributů.	211
Automatizace QTO pomocí LLM a strukturovaných dat	216
QTO výpočet celého projektu pomocí skupinových pravidel z tabulky Excelu.....	220
KAPITOLA 6.3. 4D, 6D -8D A VÝPOČET EMISÍ OXIDU UHLIČITÉHO CO₂.	226
4D model: integrace času do odhadu stavby	226
Harmonogram výstavby a jeho automatizace na základě údajů o nákladech	227
Rozšířené atributové vrstvy 6D -8D: od energetické účinnosti po zajištění bezpečnosti	229
Odhad emisí CO ₂ a výpočet emisí oxidu uhličitého ze stavebních projektů	232
KAPITOLA 6.4. STAVEBNÍ ERP A PMIS SYSTÉMY	236
Stavební ERP -systémy na příkladu výpočtů a odhadů	236
PMIS: prostředník mezi ERP a stavbou.	241
Spekulace, zisk, izolovanost a nedostatek v ERP a PMIS.....	242
Konec éry uzavřených ERP /PMIS: stavebnictví potřebuje nové přístupy.	244
Další kroky: efektivní využívání údajů o projektech.....	246
VII ČÁST CAD A BIM: MARKETING, REALITA A BUDOUCNOST PROJEKTOVÝCH DAT VE STAVEBNICTVÍ.....	249
KAPITOLA 7.1. VZNIK KONCEPCÍ BIM- VE STAVEBNICTVÍ	250
Historie vzniku BIM a otevřeného BIM jako marketingových konceptů dodavatelů CAD-.	250
Realita BIM: místo integrovaných databází - uzavřené modulární systémy.....	253
Vznik otevřeného formátu IFC ve stavebnictví	255
Problém s formátem IFC v závislosti na geometrickém jádru.....	257
Výskyt v konstrukci tématu sémantiky a ontologie	260
Proč sémantické technologie ve stavebnictví nesplňují očekávání?.....	262
KAPITOLA 7.2. UZAVŘENÉ FORMÁTY PROJEKTŮ A OTÁZKY INTEROPERABILITY	266
Uzavřená data a klesající produktivita: slepá ulička průmyslu CAD (BIM)	266
Mýtus o interoperabilitě mezi systémy CAD.....	268
Přejít na a granulární data m.....	272

KAPITOLA 7.3. GEOMETRIE VE STAVEBNICTVÍ: OD ČAR PO METRY KRYCHLOVÉ	276
Když se čáry mění v peníze aneb proč stavitelé potřebují geometrii	276
Od čar k objemům: Jak se z plochy a objemu stávají data.....	276
Přechod na MESH, USD a polygons: použití teselace pro geometrii	278
LOD, LOI, LOMD - jedinečná klasifikace detailů v CAD (BIM).....	280
Nové standardy CAD (BIM) - AIA, BEP, IDS, LOD, COBie.....	283
KAPITOLA 7.4. PARAMETRIZACE NÁVRHU A POUŽITÍ LLM PRO PROVOZ V SYSTÉMU CAD	288
Iluze jedinečnosti dat CAD (BIM): cesta k analytice a otevřeným formátům	288
Navrhování pomocí parametrů: budoucnost CAD a BIM	291
Vznik LLM v procesech zpracování dat v návrhovém CAD.....	294
Automatizovaná analýza souborů DWG pomocí LLM a Pandas	297
Další kroky: přechod od uzavřených formátů k otevřeným datům	303
VIII ČÁST ROZHODOVÁNÍ ZALOŽENÉ NA DATECH, ANALYTICE, AUTOMATIZACI A STROJOVÉM UČENÍ.	305
KAPITOLA 8.1. ANALÝZA DAT A ROZHODOVÁNÍ ZALOŽENÉ NA DATECH.	306
Data jako zdroj při rozhodování	306
Vizualizace dat: klíč k pochopení a rozhodování.....	309
KPI a ROI.....	311
Přístrojové panely a dashboardy: vizualizace metrik pro efektivní řízení.....	313
Analýza dat a umění klást otázky	315
KAPITOLA 8.2. TOK DAT BEZ RUČNÍ PRÁCE: PROČ JE POTŘEBA ETL	318
Automatizace ETL: snížení nákladů a zrychlení zpracování dat	318
ETL Výpis: sběr dat	321
ETL Transformace: použití validačních a transformačních pravidel.	324
ETL Load: Vizualizace výsledků v tabulkách a grafech.	327
ETL Load: Automatické vytváření dokumentů PDF.....	332
ETL Load: automatické generování dokumentů z FPDF.....	333
ETL Load: Reportování a načítání do jiných systémů.....	337
ETL s LLM: Vizualizace dat z PDF -dokumentů.	338
KAPITOLA 8.3. AUTOMATICKÝ ETL DOPRAVNÍK (POTRUBÍ)	343
Potrubí: Automatický dopravník ETL data.	343
Potrubí -ETL proces validace dat pomocí LLM	347
Potrubí -ETL: ověření dat a informací o prvcích projektu v CAD (BIM).....	349
KAPITOLA 8.4. ORCHESTRACE ETL A PRACOVNÍCH POSTUPŮ: PRAKTICKÁ ŘEŠENÍ.....	355

DAG a Apache Airflow: automatizace a orchestrace pracovních postupů	355
Apache Airflow: praktická aplikace automatizaci ETL.....	356
Apache NiFi pro směrování a konverzi dat	360
n8n Low-Code, No-Code orchestrace procesů.	361
Další kroky: přechod od manuálních operací k řešením založeným na analytice.....	364
IX ČÁST UKLÁDÁNÍ A SPRÁVA DAT VE STAVEBNICTVÍ.....	366
KAPITOLA 9.1. DATOVÁ INFRASTRUKTURA: OD ÚLOŽNÝCH FORMÁTŮ K DIGITÁLNÍM ÚLOŽIŠTÍM....	367
Datové atomy: základ efektivní správy informací	367
Ukládání informací: soubory nebo data.....	368
Ukládání velkých objemů dat: Analýza oblíbených formátů a jejich efektivity	370
Optimalizace úložiště pomocí aplikace Apache Parquet	373
DWH: Datový sklad datové sklady	375
Data Lake - vývoj ETL na ELT: od tradičního čištění k flexibilnímu zpracování.....	377
Architektura datových skladů: synergie skladů a datových jezer	378
CDE, PMIS, ERP nebo DWH a datové jezero.	381
KAPITOLA 9.2. SPRÁVA DATOVÉHO SKLADU A PREVENCE CHAOSU	384
Vektorové databáze a ohraničující pole.....	384
Data Governance), Data Minimalism) a Data Swamp).	387
DataOps a VectorOps: nové datové standardy.....	390
Další kroky: od chaotického úložiště ke strukturovanému úložišti.....	391
X ČÁST VELKÁ DATA, STROJOVÉ UČENÍ A PŘEDPOVĚDI.....	393
KAPITOLA 10.1. VELKÁ DATA A JEJICH ANALÝZA	394
Velká data ve stavebnictví: od intuice k předvídatelnosti	394
Zpochybnění proveditelnosti velkých dat: korelace, statistika a výběr vzorků dat	395
Velká data: analýza dat ze souboru milionů stavebních povolení v San Francisku	398
Příklad velkých dat na základě dat CAD (BIM).	403
Internet věcí a chytré smlouvy	407
KAPITOLA 10.2. STROJOVÉ UČENÍ A PŘEDPOVĚDI.....	411
Strojové učení a umělá inteligence změní způsob, jakým stavíme.....	411
Od subjektivního hodnocení ke statistické předpovědi.....	413
Titanic dataset: Hello World ve světě analytiky data a big data	415
Strojové učení v akci: od pasažérů Titaniku po řízení projektů.....	420
Předpovědi a prognózy na základě historických dat.....	424
Klíčové koncepty strojového učení	426

KAPITOLA 10.3. PŘEDPOVÍDÁNÍ NÁKLADŮ A TERMÍNU POMOCÍ STROJOVÉHO UČENÍ.....	429
Příklad použití strojového učení pro zjištění nákladů a časového harmonogramu projektu	429
Předpověď nákladů a času projektu pomocí lineární regrese	431
Předpovědi nákladů a času projektu pomocí algoritmu K-nejbližšího souseda (k-NN).....	434
Další kroky: od skladování k analýze a předpovědím.....	438
XI ČÁST STAVEBNICTVÍ V DIGITÁLNÍM VĚKU. PŘÍLEŽITOSTI A VÝZVY	441
KAPITOLA 11.1. STRATEGIE PŘEŽITÍ: BUDOVÁNÍ KONKURENČNÍCH VÝHOD.....	442
Korelace místo výpočtů: budoucnost stavební analytiky	442
Přístup založený na datech ve stavebnictví: nová úroveň infrastruktury.....	445
Digitální kancelář nové generace: jak AI mění pracovní prostor	447
Otevřená data a uberizace jsou hrozbou pro stávající stavební firmy.....	449
Nevyřešené problémy uberizace jako poslední šance využít čas k transformaci	452
KAPITOLA 11.2. PRAKTICKÝ PRŮVODCE ZAVÁDĚNÍM PŘÍSTUPU ZAŁOŽENÉHO NA DATECH	457
Od teorie k praxi: Plán digitální transformace ve stavebnictví	457
Položení digitálního základu: 1-5 kroků k digitální vyspělosti.....	459
Uvolnění potenciálu dat: 5-10 kroků k digitální zralosti	463
Plán transformace: od chaosu k datově řízené společnosti	470
Budování v průmyslu 5.0: Jak vydělávat peníze, když už se nemůžete skrývat.....	473
ZÁVĚR	475
O AUTOROVÍ	478
DALŠÍ VZTAHY	479
KOMENTÁŘ K PŘEKLADU	479
DALŠÍ DOVEDNOSTI A POJMY	480
SLOVNÍČEK	484
SEZNAM ODKAZŮ A ONLINE MATERIÁLŮ	491
INDEX PŘEDMĚTŮ	508

MAXIMÁLNÍ POHODLÍ S TIŠTĚNOU VERZÍ

Držíte v ruce bezplatnou digitální verzi knihy **Data-Driven Construction**. Pro pohodlnější práci a rychlý přístup k materiálům doporučujeme věnovat pozornost **tištěnému vydání**:



■ **Vždy po ruce:** kniha v tištěné podobě se stane spolehlivým pracovním nástrojem, který vám umožní rychle najít a použít správné vizualizace a diagramy v jakékoli pracovní situaci.

■ **Vysoké kvalitní ilustrace:** všechny obrázky a grafika v tištěném vydání jsou prezentovány v maximální kvalitě.

■ **Rychlý přístup k informacím:** snadná navigace, možnost dělat si poznámky, záložky a pracovat s knihou kdekoli.

Zakoupením plné tištěné verze knihy získáte praktický nástroj pro pohodlnou a efektivní práci s informacemi: možnost rychle využívat obrazové materiály při každodenních úkolech, rychle vyhledávat potřebná schéma a dělat si poznámky. Navíc svým nákupem podpoříte šíření otevřených znalostí.

Tištěnou verzi knihy si můžete objednat na adrese: datadrivenconstruction.io/books



I ČÁST

OD HLINĚNÝCH TABULEK K DIGITÁLNÍ REVOLUCI: JAK SE VYVÍJELY INFORMACE VE STAVEBNICTVÍ

První část knihy se zabývá historickým vývojem správy dat ve stavebnictví od primitivních záznamů na fyzických nosičích až po moderní digitální ekosystémy. Analyzuje transformaci technologií pro správu informací, vznik ERP -systémů a dopad fragmentace dat na efektivitu podnikových procesů. Zvláštní pozornost je věnována procesu digitalizace informací a rostoucímu významu objektivní analýzy namísto subjektivního expertního posouzení. Podrobně je zkoumán exponenciální růst objemu informací, kterému čelí moderní stavebnictví, a s tím spojené výzvy pro podnikové systémy. Zkoumá se postavení stavebnictví v kontextu čtvrté a páté průmyslové revoluce, jakož i potenciál využití umělé inteligence a datově orientovaných přístupů k vytvoření udržitelné konkurenční výhody.

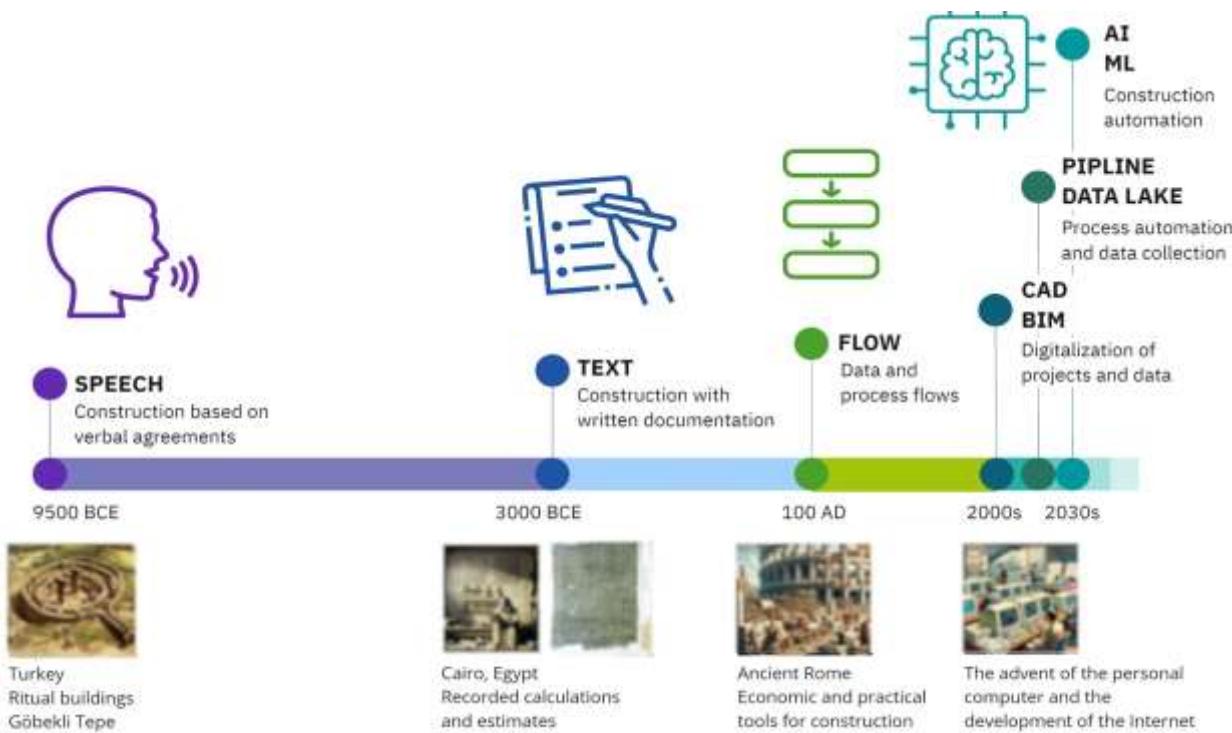
KAPITOLA 1.1.

VÝVOJ VYUŽÍVÁNÍ DAT VE STAVEBNICTVÍ

Zrod datové éry ve stavebnictví

Přibližně před 10 000 lety, v období neolitu, došlo k převratné změně ve vývoji lidstva, které opustilo kočovný způsob života a začalo žít usedle, což vedlo ke vzniku prvních primitivních staveb z hliny, dřeva a kamene [6]. Od tohoto okamžiku začínají dějiny stavebnictví.

S rozvojem civilizací se architektura stávala stále složitější a vznikaly první rituální chrámy a veřejné budovy. Rostoucí složitost architektonických návrhů vyžadovala, aby inženýři a manažeři ve starověku vytvořili první záznamy a výpočty. První záznamy na hliněných tabulkách a papyrech často obsahovaly popis logiky výpočtu množství potřebných stavebních materiálů, jejich ceny a výpočtu platby za provedenou práci [7]. Tak začala éra využívání dat ve stavebnictví - dlouho před nástupem moderních digitálních technologií (obrázek 1.1-1).

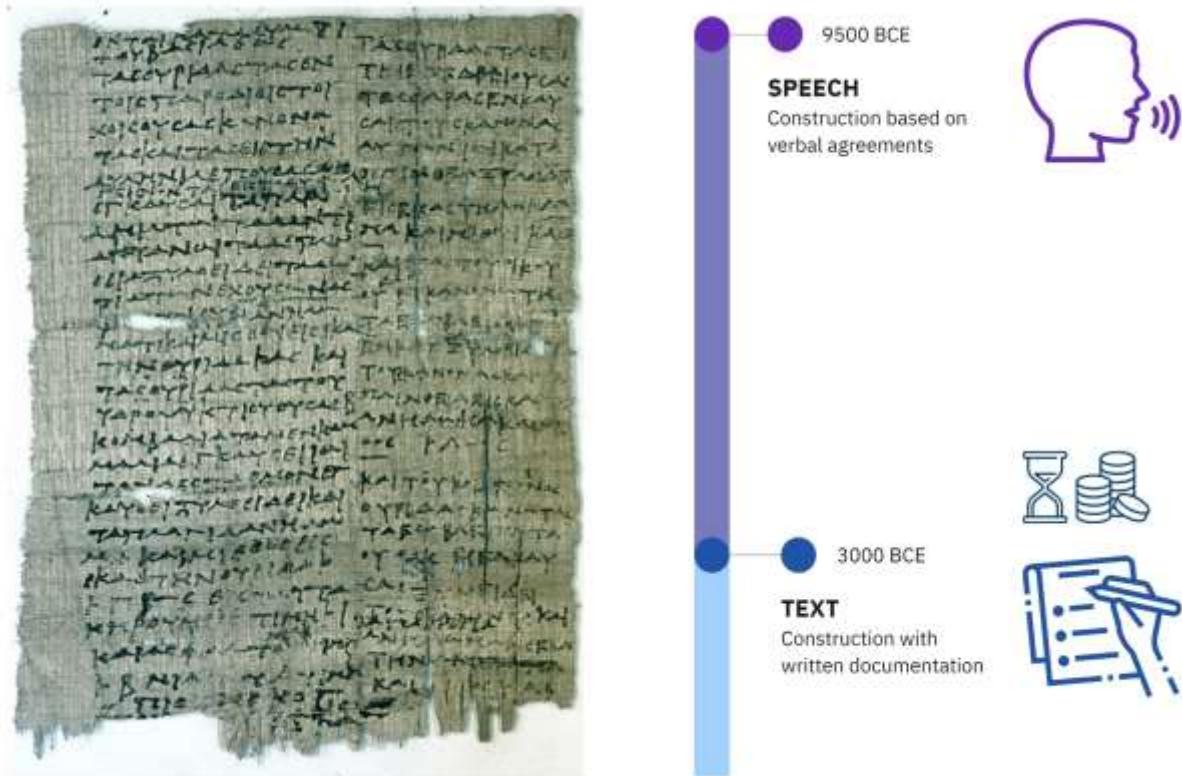


Obr. 1.1-1 Chronologie vývoje informačních technologií ve stavebnictví: od verbálních informací k umělé inteligenci.

Od hlíny a papyru k digitální technologii

První doklady o stavbě pocházejí z období stavby pyramid, tedy z doby kolem 3000-4000 let př. n. l.[7]. Od té doby vedení písemných záznamů usnadňovalo a doprovázelo pokrok ve stavebnictví a umožňovalo hromadění a systematizaci poznatků, které v průběhu následujících 10 000 let vedly k významným inovacím ve stavebních metodách a architektuře.

Použití prvních fyzických nosičů ve stavebnictví, jako jsou hliněné tabulky, papyrus z doby před tisíci lety (obrázek 1.1-2) nebo papír "A0" v 80. letech 20. století, k záznamu dat nebylo původně určeno k použití těchto informací pro nové projekty. Hlavním účelem těchto záznamů bylo podrobně popsat aktuální stav projektu, včetně výpočtů potřebných materiálů a nákladů na práci. Stejně tak v dnešním světě dostupnost digitálních projektových dat a modelů nezaručuje vždy jejich použití v budoucích projektech a často slouží především jako informace pro aktuální výpočty potřebných materiálů a nákladů na stavbu.



Obr. 1.1-2 Papyrus ze 3. století př. n. l. popisující náklady na malbu různých typů oken v královském paláci technikou enkaustiky.

Lidstvu trvalo přibližně 5 000 let, než přešlo od ústních rozhovorů k písemným dokumentům v oblasti řízení stavebních projektů, a stejnou dobu trvalo, než přešlo od papíru k digitálním datům jako primárnímu zdroji pro plánování a řízení.

Stejně jako rozvoj obchodních a peněžních vztahů podnítil vznik písma a prvních právníků pro řešení sporů, tak i první záznamy o materiálových nákladech a rozsahu prací ve stavebnictví vedly ke vzniku prvních manažerů ve stavebnictví, jejichž povinností bylo dokumentovat, sledovat a odpovídat za klíčové informace o časovém plánu projektu a nákladech.

Dnes hrají data mnohem významnější roli: nejenže zaznamenávají učiněná rozhodnutí, ale stávají se také nástrojem pro předvídání a modelování budoucnosti. To je základ, na kterém je postaven moderní procesní přístup v projektovém řízení - proměňuje nashromážděné zkušenosti v systém rozhodování založený na strukturovaných a ověřitelných datech.

Proces jako nástroj pro zkušenosti založené na datech

Jádrem každého procesu je přeměna minulých zkušeností na nástroj pro plánování budoucnosti. Zkušenost v moderním slova smyslu je strukturovaný soubor dat, který lze analyzovat a na jeho základě provádět informované předpovědi.

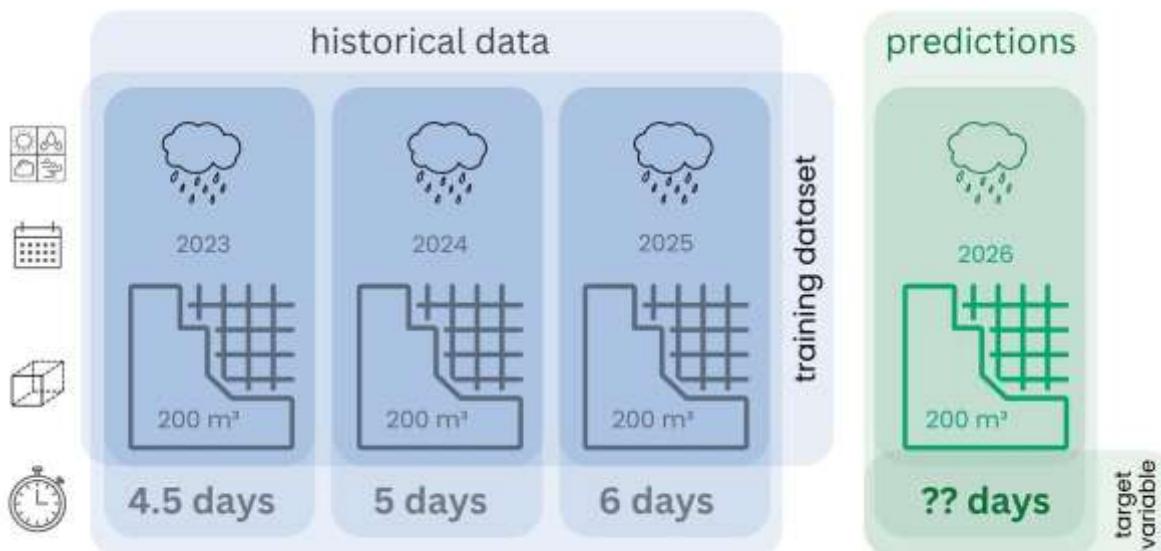
Základem prognózování jsou historická data, která jasně ukazují výsledky provedené práce a umožňují nahlédnout do faktorů, které tyto výsledky ovlivňují.

Uveďme si konkrétní příklad z monolitické stavby: při plánování harmonogramu prací se obvykle zohledňuje objem betonu, složitost konstrukce a povětrnostní podmínky. Předpokládejme, že konkrétní stavbyvedoucí nebo historické údaje firmy za poslední tři roky (2023-2025) ukazují, že vylití monolitické konstrukce o ploše 200 m² za deštivého počasí trvalo 4,5 až 6 dní (obr. 1.1-3). Právě tyto nashromážděné statistiky se stávají základem pro předvídání doby realizace a kalkulaci nákladů na zdroje při plánování podobných prací v budoucích projektech. Na základě těchto historických údajů může mistr nebo odhadce na základě zkušeností kvalifikovaně předpovědět dobu potřebnou k dokončení budoucích podobných prací v roce 2026 za podobných podmínek.

V tomto případě odhadu času funguje analytický proces jako mechanismus pro transformaci různorodých dat do strukturovaných zkušeností a následně do přesného plánovacího nástroje. Data a procesy tvoří jeden ekosystém, kde jedno bez druhého nemůže existovat.

Počítejte, co se dá spočítat, měřte, co se dá změřit, a to, co se nedá změřit, udělejte měřitelným [8].

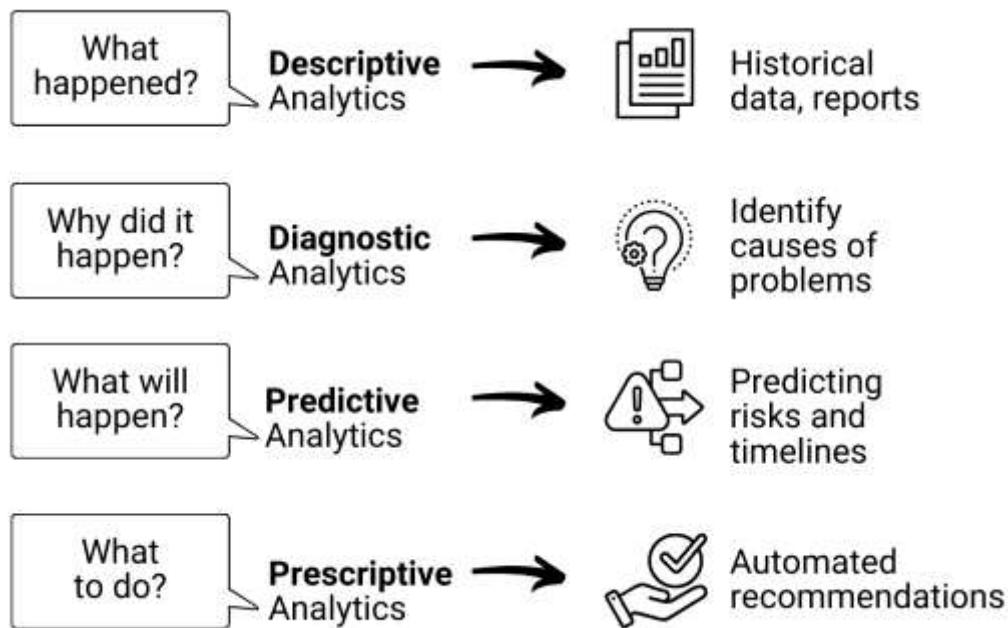
- Galileo Galilei



Obr. 1.1-3 Historická data slouží jako trénovací soubor dat pro předpověď jedné z hodnot v budoucnosti.

V dnešním podnikatelském prostředí se analýza dat stává důležitou součástí efektivního řízení projektů, optimalizace procesů a strategického rozhodování. Stavebnictví si postupně osvojuje čtyři klíčové úrovně analytiky, z nichž každá odpovídá na konkrétní otázku a přináší jedinečné výhody (obr. 1.1-4):

- **Deskriptivní analýza** - odpovídá na otázku "co se stalo?" a poskytuje historická data a zprávy o minulých událostech a výsledcích: za poslední tři roky (2023-2025) trvalo vylití monolitické konstrukce o ploše 200 m² za deštivého počasí 4,5 až 6 dní.
- **Diagnostická analýza** - odpovídá na otázku "proč se to stalo?" tím, že identifikuje příčiny problémů: analýza ukazuje, že doba lití monolitické konstrukce se prodloužila v důsledku deštivého počasí, které zpomalilo proces vytvrzování betonu.
- **Prediktivní analýza** - výhled do budoucna, předpověď možných rizik a doby realizace na základě odpovědi na otázku "co se stane?": na základě historických údajů se předpokládá, že vylití podobné monolitické konstrukce o rozloze 200 m² za deštivého počasí v roce 2026 bude trvat přibližně 5,5 dne, přičemž se zohlední všechny známé faktory a trendy.
- **Preskriptivní analýza** - poskytuje automatizovaná doporučení a odpovědi na otázku "co dělat?" a umožňuje firmám zvolit optimální kroky: Pro optimalizaci prací se například doporučuje: použít speciální přísady pro urychlení tvrdnutí betonu v podmínkách vysoké vlhkosti; naplánovat betonáž na období s nejnižší pravděpodobností srážek; zajistit dočasné přístřešky pro konstrukci, které zkrátí dobu prací na 4-4,5 dne i v nepříznivých povětrnostních podmínkách.



Obr. 1.1-4 Hlavní typy analytiky: od popisu minulosti po automatizované rozhodování.

Plnohodnotná digitální transformace, která znamená přechod na systémovou analytiku a řízení založené na datech, vyžaduje nejen outsourcing, ale i vytvoření kompetentního interního týmu. Klíčovými členy takového týmu by měli být produktoví manažeři, datoví inženýři, analytici a vývojáři, kteří budou úzce spolupracovat s obchodními jednotkami (obr. 4.3-9). Tato spolupráce je nezbytná pro kladení inteligentních analytických otázek a efektivní parametrizaci obchodních rozhodovacích úloh. V informační společnosti se data stávají nejen pomocným nástrojem, ale základem pro prognózování a optimalizaci.

Digitální transformace ve stavebnictví zásadním způsobem mění způsob navrhování, správy a provozu zařízení. Tento proces se označuje jako digitalizace informací - kdy se všechny aspekty stavebního procesu digitalizují do digitální podoby vhodné pro analýzu.

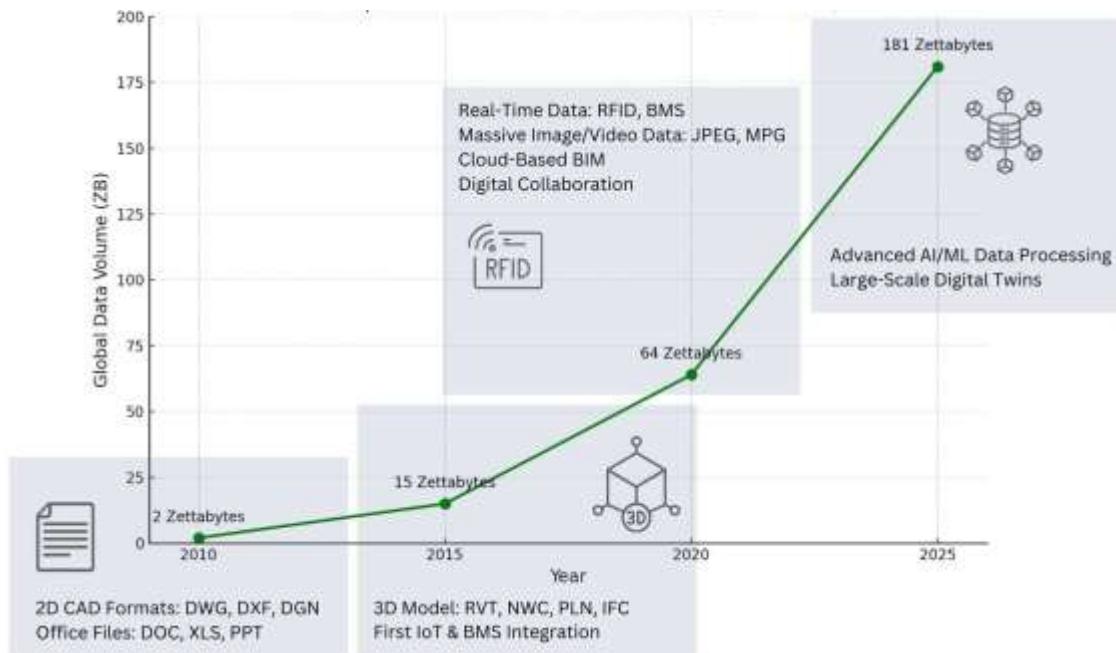
Digitalizace informací o stavebním procesu

Po tisíciletí se množství informací zaznamenaných ve stavebnictví téměř nezměnilo, ale v posledních desetiletích rychle roste (obrázek 1.1-5).

Podle studie společnosti PwC® "Managed Data. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9] bylo 90% všech dat na světě vytvořeno v posledních dvou letech (k roku 2015). Většina společností však tato data plně nevyužívá, protože buď zůstávají v oddělených systémech, nebo jsou jednoduše archivována bez skutečné analýzy.

Nárůst objemu dat se v posledních letech jen zrychlil a zdvojnásobil se z 15 zettabytů v roce 2015 na

181 zettabytů v roce 2025 [10]. Každý den se servery stavebních a projekčních firem plní projektovou dokumentací, harmonogramy prací, výpočty a kalkulacemi, finančními zprávami. Pro 2D/3D -výkresy se používají formáty DWG, DXF a DGN a pro 3D modely - RVT, NWC, PLN a IFC™. Textové dokumenty, tabulky a prezentace se ukládají ve formátech DOC, XLSX a PPT. Kromě videí a snímků ze stavby - ve formátech MPG a JPEG, jsou v reálném čase ukládána data z komponent IoT, tagů RFID® (identifikace a sledování) a systémů řízení budov BMS (monitorování a řízení)



Obr. 1.1-5 Parabolický růst dat v letech 2010-2025 (podle [10]).

S rychlým nárůstem informací se stavebnictví potýká s potřebou nejen shromažďovat a uchovávat data, ale také zajistit jejich ověřování, validaci, měřitelnost a analytické zpracování. V současné době prochází odvětví aktivní fází digitalizace informací - systematické transformace všech aspektů stavební činnosti do digitální podoby vhodné pro analýzu, interpretaci a automatizaci.

Digitalizace informací znamená převzít informace o všech subjektech a prvcích stavebního projektu a samotného stavebního procesu - včetně těch, které jsme dříve za informace vůbec nepovažovali - a převést je do datového formátu, aby byly informace kvantifikovatelné a snadno analyzovatelné.

V kontextu stavebnictví to znamená zachycovat a digitalizovat informace všech prvcích projektů a všech procesech - od pohybu strojů a lidí na staveništi až po povětrnostní a klimatické podmínky na staveništi, aktuální ceny materiálů a úrokové sazby centrálních bank - s cílem vytvářet analytické modely.

Pokud dokážete změřit to, o čem mluvíte, a vyjádřit to v číslech, pak o daném tématu něco víte. Pokud to však nedokážete vyjádřit kvantitativně, jsou vaše znalosti velmi omezené a neuspokojivé. Může to být výchozí bod, ale není to úroveň skutečného vědeckého poznání. [11].

- W. Thomson (lord Kelvin), 1824-1907, britský vědec

Digitalizace informací dalece přesahuje tradiční přístup ke sběru informací, kdy se zaznamenávaly pouze základní ukazatele, jako jsou člověkohodiny nebo skutečné náklady na materiál. Dnes lze prakticky jakoukoli událost transformovat do proudu dat vhodných pro hloubkovou analýzu pomocí pokročilých analytických nástrojů a technik strojového učení. Stavebnictví prošlo zásadním posunem od papírových výkresů, excelových tabulek a ústních pokynů k digitálním systémům (obrázek 1.2-4), v nichž se každý prvek projektu stává zdrojem dat. Dokonce i zaměstnanci - od inženýrů až po stavební dělníky na stavbě - jsou nyní vnímáni jako soubor digitálních proměnných a datových souborů.

Podle studie společnosti KPMG "Familiar Challenges - New Approaches: Global Construction Survey 2023" se digitální dvojčata, umělá inteligence (AI) a velká data, stávají klíčovými faktory pro zlepšení ziskovosti projektů [2].

Moderní technologie nejenž zjednoduší sběr informací a do značné míry jej automatizují, ale také radikálně snižují náklady na ukládání dat. V důsledku toho podniky upouštějí od selektivního přístupu a dávají přednost ukládání celé řady informací pro pozdější analýzu (obr. 2.1-5), což otevírá potenciální možnosti optimalizace procesů v budoucnosti.

Digitalizace informací a digitalizace umožňují odhalit skrytu, dosud nevyužitou hodnotu informací. Pokud jsou data správně uspořádána, lze je znova použít, znova interpretovat a integrovat do nových služeb a řešení.

V budoucnu digitalizace informací pravděpodobně povede k plné automatizaci správy dokumentů, zavedení samosprávných stavebních procesů a vzniku nových profesí - stavebních datových analytiků, odborníků na řízení projektů s umělou inteligencí a digitálních inženýrů. Stavební projekty se stanou dynamickými zdroji informací a rozhodování nebude založeno na intuici nebo subjektivní zkušenosti, ale na spolehlivých a reprodukovatelných digitálních faktech

Informace jsou ropou 21. století a analytika je spalovacím motorem [12].

- Peter Sondergaard, senior viceprezident, Gartner®

Podle studie IoT Analytics 2024 [13] se očekává, že celosvětové výdaje na správu a analýzu dat dramaticky vzrostou ze 185,5 miliardy dolarů v roce 2023 na 513,3 miliardy dolarů do roku 2030, což představuje 16% roční míru růstu. Ne všechny komponenty však rostou stejným tempem: rychle roste analytika, zatímco růst úložišť se zpomaluje. Analytika zajistí nejrychlejší růst v ekosystému správy dat: předpokládá se, že vzroste z 60,6 miliardy USD v roce 2023 na 227,9 miliardy USD do roku 2030, což představuje složenou roční míru růstu 27%.

Se zrychlující se digitalizací informací a rychlým růstem jejich objemu čelí management stavebních projektů a společností potřebě systematicky ukládat, analyzovat a zpracovávat různorodá, často heterogenní data. V reakci na tuto výzvu začal od poloviny 90. let minulého století v oboru masivní přechod na elektronickou tvorbu, ukládání a správu dokumentace - od tabulek a projektových výpočtů až po výkresy a smlouvy.

Tradiční papírové dokumenty, které vyžadují podpisy, fyzické ukládání, pravidelné revize a archivaci v kartotékách, jsou postupně nahrazovány digitálními systémy, které ukládají data strukturovaným způsobem - v databázích specializovaných aplikací.



KAPITOLA 1.2.

TECHNOLOGIE A SYSTÉMY ŘÍZENÍ V MODERNÍM STAVEBNICTVÍ

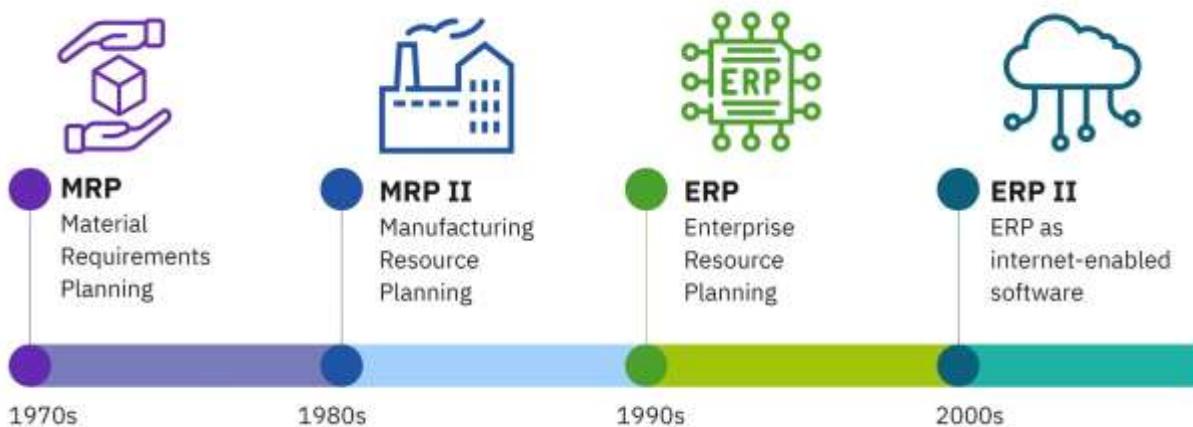
Digitální revoluce a vznik modulárních MRP/ERP -systémů

Éra moderního digitálního ukládání a zpracování dat začala s příchodem magnetické pásky v 50. letech 20. století, která otevřela možnost ukládání a využívání velkého množství informací. Dalším průlomem byl nástup diskových mechanik, který radikálně změnil přístup ke správě dat ve stavebnictví.

S rozvojem datových skladů vstoupilo na trh řešení velké množství společností, které začaly vyvíjet modulární software pro vytváření, ukládání a zpracování dat a automatizaci rutinních úloh

Exponenciální nárůst informací a nástrojů vedl k potřebě integrovaných modulárních řešení, která nepracují s jednotlivými soubory, ale pomáhají spravovat a řídit tok dat napříč procesy a projekty.

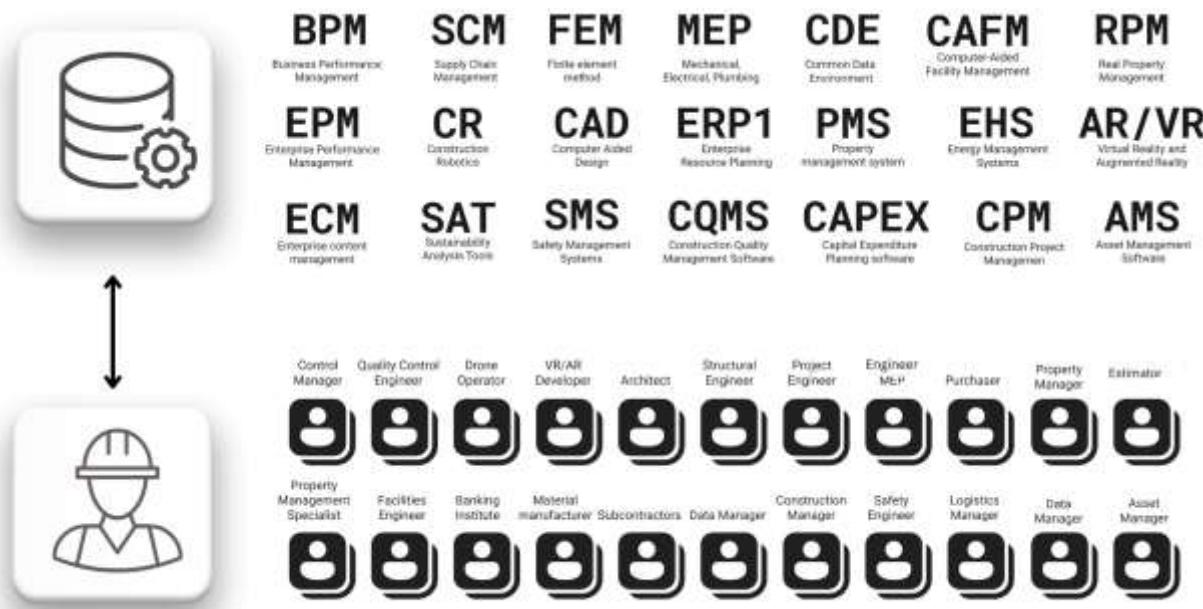
První komplexní nástroje platformy musely nejen ukládat dokumenty, ale také dokumentovat všechny požadavky na změny a operace v procesech: kdo je inicioval, jaký byl rozsah požadavku a co bylo nakonec zaznamenáno jako hodnota nebo atribut. Pro tyto účely byl zapotřebí systém, který by dokázal sledovat přesné výpočty a učiněná rozhodnutí (obr. 1.2-1). Takovými platformami byly první systémy MRP (Material Requirements Planning) a ERP (Enterprise Resource Planning), které se staly populárními od počátku 90. let [14]



Obrázek 1.2-1 Pokrok v technologii ukládání dat vedl v 80. letech 20. století ke vzniku ERP - systémů.

První systémy MRP - a ERP - položily základ éry digitalizace v oblasti řízení obchodních procesů a stavebních projektů. Modulární systémy, původně určené k automatizaci klíčových podnikových procesů, byly nakonec integrovány s dalšími, flexibilnějšími a přizpůsobivějšími softwarovými řešeními.

Tato doplňková řešení byla určena pro zpracování dat a správu obsahu projektů (obr. 1.2-2), buď nahrazovala některé moduly velkých systémů, nebo je účinně doplňovala a rozšiřovala funkčnost celého systému.



Obr. 1.2-2 Nová softwarová řešení přilákala do podniku armádu manažerů, kteří mají za úkol řídit datové toky.

V posledních desetiletích společnosti investovaly velké prostředky do modulárních systémů [15] a vnímaly je jako dlouhodobá integrovaná řešení.

Podle zprávy Software Path pro rok 2022 [16] činí průměrný rozpočet na jednoho uživatele ERP -systému 9 000 USD. V průměru tyto systémy využívá asi 26% zaměstnanců společnosti. Pro organizaci se 100 uživateli tak celkové náklady na implementaci ERP dosahují přibližně 900 000 USD.

Investice do proprietárních, uzavřených a modulárních řešení jsou stále méně opodstatněné na pozadí rychlého rozvoje moderních, flexibilních a otevřených technologií. Pokud již byly takové investice provedeny, je důležité objektivně přehodnotit roli stávajících systémů: zda jsou i nadále dlouhodobě nezbytné, nebo zda lze jejich funkce revidovat a realizovat efektivněji a transparentněji.

Jedním z klíčových problémů dnešních modulárních datových platform je, že centralizují správu dat v rámci uzavřených aplikací. V důsledku toho se data - hlavní aktivum společnosti - stávají závislými na konkrétních softwarových řešeních, a nikoli naopak. To omezuje opakování použití informací, komplikuje migraci na a snižuje agilitu podniku v rychle se měnícím digitálním prostředí.

Pokud je pravděpodobné, že význam či relevance uzavřené modulární architektury v budoucnosti poklesne, má smysl uznat dnes vynaložené náklady jako utopené náklady a zaměřit se na strategický přechod k otevřenějšímu, škálovatelnějšímu a přizpůsobivějšímu digitálnímu ekosystému.

Proprietární software se vyznačuje výhradní kontrolou vývojáře nad zdrojovým kódem a uživatelskými daty vytvořenými v rámci používání těchto řešení. Na rozdíl od softwaru s otevřeným zdrojovým kódem nemají uživatelé přístup k vnitřní struktuře aplikace a nemohou ji nezávisle revidovat, upravovat nebo přizpůsobovat svým potřebám. Místo toho jsou povinni zakoupit licence, které jim poskytují právo používat software v mezích stanovených dodavatelem.

Moderní přístup zaměřený na data nabízí jiné paradigma: data by měla být vnímána jako hlavní strategické aktivum - nezávislá, trvalá a oddělená od konkrétních softwarových řešení. Aplikace se naopak stávají pouhými datovými nástroji, které lze libovolně nahrazovat bez rizika ztráty kritických informací.

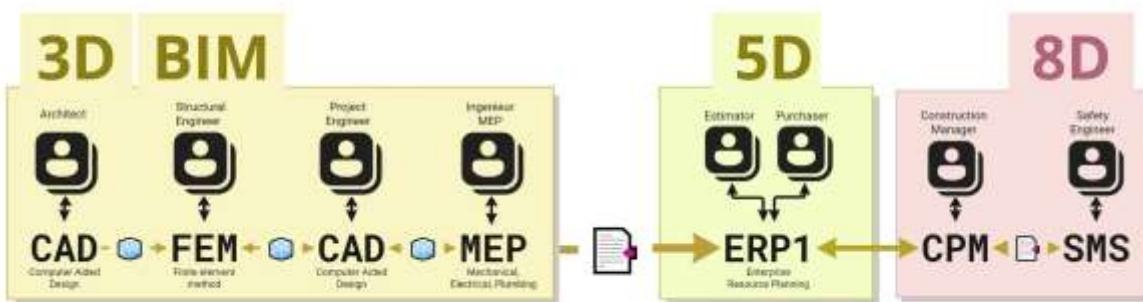
Vývoj systémů ERP a MRP v 90. letech 20. století (obrázek 1.2-1) poskytl podnikům výkonné nástroje pro řízení procesů, ale měl také nezamýšlený důsledek v podobě výrazného zvýšení počtu lidí, kteří se podílejí na udržování informačních toků. Namísto automatizace a zjednodušení provozních úkolů tyto systémy často vytvořily novou úroveň složitosti, byrokracie a závislosti na interních IT zdrojích.

Systémy pro správu dat: od dolování dat k obchodním výzvám

Dnešní společnosti se potýkají s nutností integrovat více systémů pro správu dat. Výběr systémů pro správu dat, rozumná správa těchto systémů a integrace různorodých zdrojů dat se stávají rozhodujícími pro výkonnost podniku.

V polovině 20. let 20. století můžete najít stovky (a ve velkých stavebních firmách tisíce) různých systémů (obrázek 1.2-3), které musí pracovat v souladu, aby všechny aspekty stavebního procesu probíhaly hladce a soudržně.

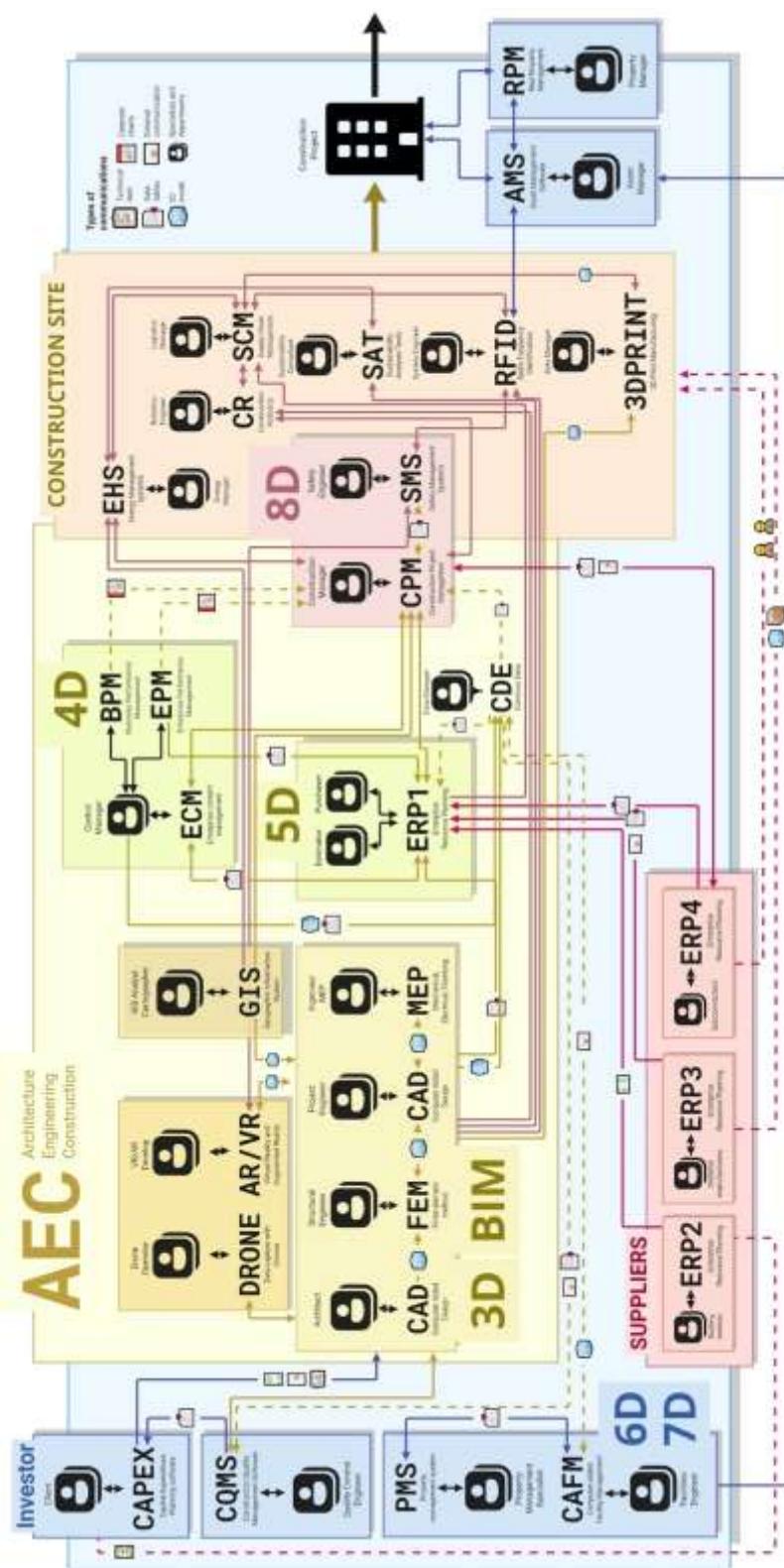
Podle studie společnosti Deloitte z roku 2016® "Data-Driven Management in Digital Capital Projects" - průměrný stavební odborník používá denně 3,3 softwarových aplikací, ale pouze 1,7 z nich je vzájemně integrováno [17].



Obr. 1.2-3 Každý podnikový systém vyžaduje profesionální tým a odpovědného manažera pro kvalitní správu dat.

Níže je uveden seznam oblíbených systémů pro střední a velké společnosti ve stavebnictví, které se používají při efektivním řízení stavebních projektů:

- **ERP (Enterprise Resource Planning)** - zajišťuje integraci podnikových procesů včetně účetnictví, nákupu a řízení projektů.
- **CAPEX (software pro plánování kapitálových výdajů)** - slouží k rozpočtování a řízení finančních investic do stavebních projektů, pomáhá stanovit náklady na dlouhodobý majetek a investice do dlouhodobého majetku.
- **CAD (Computer-Aided Design) a BIM (Building Information Modeling)** - slouží k vytváření podrobných a přesných technických výkresů a 3D -modelů projektů. Tyto systémy se zaměřují na práci s geometrickými informacemi.
- **MEP (Mechanical, Electrical, Plumbing)** - inženýrské systémy, které zahrnují mechanické, elektrické a vodovodní komponenty a podrobně popisují vnitřní "oběhový" systém projektu.
- **GIS (geografické informační systémy)** - slouží k analýze a plánování území, včetně kartografie a prostorové analýzy.
- **CQMS (software pro řízení kvality ve stavebnictví)** - zajišťuje, aby stavební procesy byly v souladu se stanovenými normami a předpisy, a pomáhá tak eliminovat nedostatky.
- **CPM (řízení stavebních projektů)** - zahrnuje plánování, koordinaci a řízení stavebních procesů.
- **CAFM (Computer-Aided Facility Management)** - systémy pro správu a údržbu budov.
- **Řízení dodavatelského řetězce (SCM)** je nutné k optimalizaci toku materiálů a informací mezi dodavateli a staveništěm.
- **EPM (Enterprise Performance Management)** - zaměřený na zlepšování podnikových procesů a výkonnosti.
- **AMS (Software pro správu majetku)** - slouží k optimalizaci používání, správy a údržby zařízení a infrastruktury po celou dobu životnosti majetku.
- **RPM (Správa nemovitostí)** - zahrnuje úkoly a procesy související se správou a provozem budov a pozemků, jakož i souvisejících zdrojů a majetku.



Obr. 1.2-4 Propojení systémů, které propojuje procesy ve firmě s tokem informací mezi jednotlivými odděleními.

- **CAE (Computer-Aided Engineering)** - Počítačově podporované inženýrství zahrnuje výpočetní a simulační systémy, jako je analýza konečných prvků (FEA) a výpočetní dynamika tekutin (CFD).
- **CFD (Computational Fluid Dynamics)** - výpočetní dynamika tekutin, modelování proudění tekutin a plynů. Podkategorie CAE.
- **CAPP (Computer-Aided Process Planning)** - Počítačem podporované plánování procesů. Slouží k vytváření map tras a procesů.
- **CAM (Computer-Aided Manufacturing)** - počítačem podporovaná výroba, tvorba řídicích programů pro CNC stroje.
- **PDM (Product Data Management)** - Správa dat o výrobku, systém pro ukládání a správu technické dokumentace.
- **MES (Manufacturing Execution System)** je systém řízení výrobních procesů v reálném čase.
- **PLM (Product Lifecycle Management)** - řízení životního cyklu prvku projektu, integruje PDM, CAPP, CAM a další systémy pro kompletní řízení výrobku od vývoje až po likvidaci.

Tyto a mnohé další systémy, zahrnující různá softwarová řešení, se staly nedílnou součástí moderního stavebnictví (obr. 1.2-4). Podstatou těchto systémů jsou specializované databáze s intuitivním rozhraním, které umožňují efektivní zadávání, zpracování a analýzu informací ve všech fázích projektování a výstavby. Vzájemná integrace digitálních nástrojů pomáhá nejen optimalizovat pracovní procesy, ale také výrazně zvyšuje přesnost rozhodování, což má pozitivní vliv na časovou náročnost a kvalitu realizace projektu.

V polovině případů však k integraci nedochází. Podle statistik je pouze každá druhá aplikace nebo systém integrován s jinými řešeními [17]. To svědčí o pokračující roztríštěnosti digitálního prostředí a zdůrazňuje potřebu vyvinout otevřené standardy a jednotná rozhraní, která zajistí komplexní výměnu informací v rámci stavebního projektu.

Jednou z hlavních výzev v oblasti integrace pro moderní společnosti zůstává vysoká složitost digitálních systémů a požadavky na uživatelské kompetence nezbytné pro efektivní vyhledávání a interpretaci informací. Pro podporu každého systému implementovaného v podniku je vytvořen tým specialistů v čele s klíčovým manažerem (obr. 1.2-2).

Klíčový správce systému hraje rozhodující roli ve správném směru toku dat a je zodpovědný za kvalitu výsledných informací, stejně jako byli první správci před tisíci lety zodpovědní za čísla zapsaná na papyru nebo hliněných tabulkách.

Aby bylo možné proměnit nesourodé informační toky v nástroj řízení, je nezbytná schopnost systematicky integrovat a spravovat data. V této architektuře musí manažeři vystupovat jako prvky jednotné sítě - jako mycelium, které spojuje jednotlivé části podniku do souvislého živého organismu schopného se přizpůsobovat a vyvíjet.

Firemní mycelium: jak se data propojují s podnikovými procesy

Proces integrace dat do aplikací a databází je založen na agregaci informací z různých zdrojů, včetně různých oddělení a specialistů (obrázek 1.2-4). Specialisté vyhledávají relevantní data, zpracovávají je

a přenášejí do svých systémů a aplikací k dalšímu využití.

Každý podnikový systém, tvořený souborem nástrojů, technologií a databází, je stromem znalostí, který je zakořeněn v půdě historických dat a roste, aby přinesl nové plody v podobě hotových řešení: dokumentů, výpočtů, tabulek, grafů a informačních panelů (obr. 1.2-5). Systémy v podniku, podobně jako stromy v určitém kousku lesa, spolu vzájemně komunikují a tvoří komplexní a dobře strukturovaný systém podporovaný a řízený odbornými manažery.

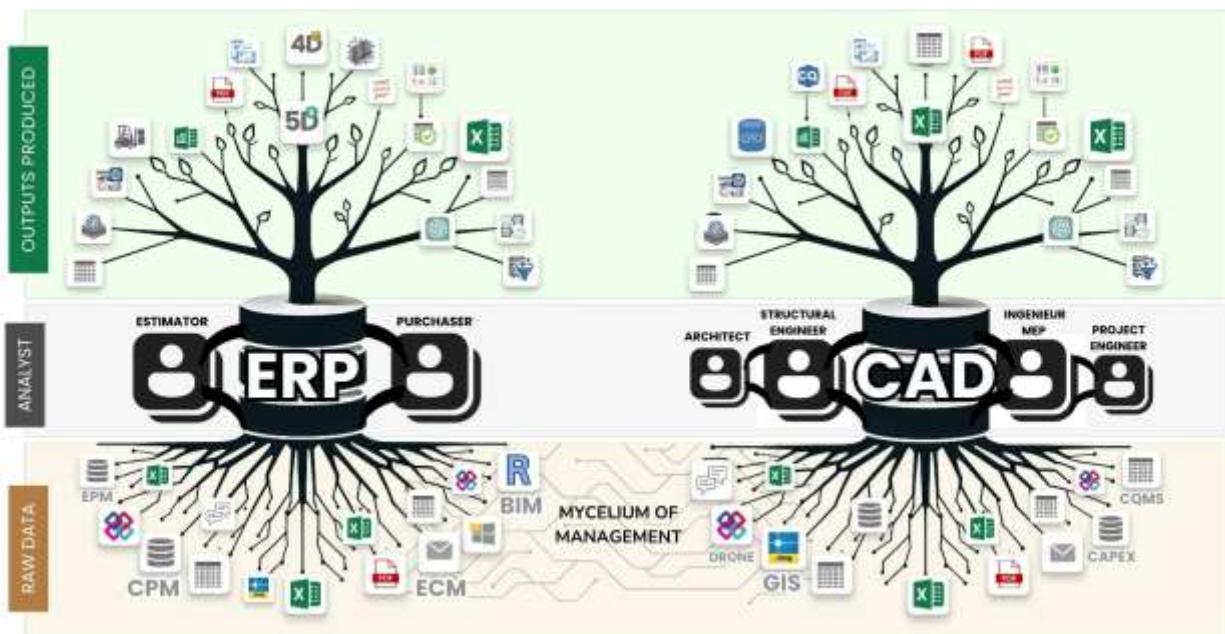
Systém získávání a přenosu informací v podniku funguje jako složitá lesní síť složená ze stromů (systémů) a podhoubí (manažerů), které fungují jako vodiče a recyklátory a zajišťují přenos informací a jejich tok do správných systémů. To pomáhá udržovat zdravý a efektivní tok a distribuci dat v rámci společnosti.

Odborníci, stejně jako kořeny, absorbuje surová data v počátečních fázích projektu a mění je na živiny pro podnikový ekosystém. Systémy pro správu dat a obsahu (obr. 1.2-4 - ERP, CPM, BIM, atd.) fungují jako výkonné informační dálnice, kterými tyto znalosti kolují po všech úrovních společnosti.

Stejně jako v přírodě, kde každý prvek ekosystému hraje svou roli, i v podnikovém prostředí přispívá každý účastník procesu - od inženýra po analytika - k růstu a plodnosti informačního prostředí. Tyto systémové "datové stromy" (obr. 1.2-5) nepředstavují jen mechanismus pro shromažďování informací, ale konkurenční výhodu, která zajišťuje udržitelnost podniku.

Lesní ekosystémy jsou překvapivě přesným odrazem toho, jak jsou organizovány digitální podnikové struktury. Stejně jako stupňovitá struktura lesa - od podrostu až po korunu stromů - přiřazuje řízení podniku úkoly k úrovním odpovědnosti a funkčním oddělením.

Hluboké a rozvětvené kořeny stromů zajišťují odolnost a přístup k živinám. Podobně pevná organizační struktura a stabilní procesy pro práci s kvalitními daty podporují celý informační ekosystém společnosti a přispívají k jejímu udržitelnému růstu a rozvoji i v obdobích nestability a krizí (silného větru) na trhu.



Obr. 1.2-5 Integrace dat prostřednictvím různých systémů je jako podhoubí, které spojuje manažery a specialisty do jednotné informační sítě.

Moderní chápání rozsahu v podnikání se vyvinulo. Hodnotu podniku dnes neurčuje pouze jeho viditelná část - "korunku" v podobě závěrečných dokumentů a zpráv - ale také hloubka "kořenového systému" kvalitativně shromážděných a systematicky zpracovaných dat. Čím více informací se podaří shromáždit a zpracovat, tím vyšší je hodnota podniku. Společnosti, které metodicky hromadí "kompost" již zpracovaných dat a jsou schopny z něj vytěžit užitečné poznatky, získávají strategickou výhodu

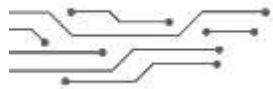
Historické informace se stávají novým druhem kapitálu, který umožňuje růst, optimalizaci procesů a konkurenční výhodu. Ve světě založeném na datech nevyhrává ten, kdo jich má více, ale ten, kdo jich více zná.

Pro stavebnictví to znamená přechod na řízení projektů v reálném čase, kdy budou všechny procesy - od návrhu a zadávání zakázek až po koordinaci dodavatelů - založeny na relevantních, denně aktualizovaných datech. Integrace informací z různých zdrojů (ERP -systémy, CAD -modely, senzory IoT na stavbách, RFID) umožní vytvářet přesnější prognózy, rychle reagovat na změny a vyhnut se zpožděním způsobeným nedostatkem aktuálních dat.

Podle studie společnosti McKinsey & Company Data-Driven Enterprise 2025 (McKinsey & Company®, 2022 [18]) budou úspěšné společnosti budoucnosti spoléhat na data ve všech klíčových aspektech své činnosti, od strategických rozhodnutí až po provozní interakce.

Data přestanou být pouhým analytickým nástrojem a stanou se nedílnou součástí všech podnikových procesů, zajistí transparentnost, kontrolu a automatizaci řízení. Datově řízený krok podle umožní organizacím minimalizovat vliv lidského faktoru, snížit provozní rizika a zvýšit transparentnost a efektivitu rozhodování.

21. století převrací ekonomické paradigma naruby: zatímco dříve se ropě říkalo "černé zlato" kvůli její schopnosti pohánět stroje a dopravu, dnes se historická data, stlačená pod časovým tlakem, stávají novým strategickým zdrojem, který pohání nikoli stroje, ale rozhodovací algoritmy, jež budou řídit podnikání.



KAPITOLA 1.3.

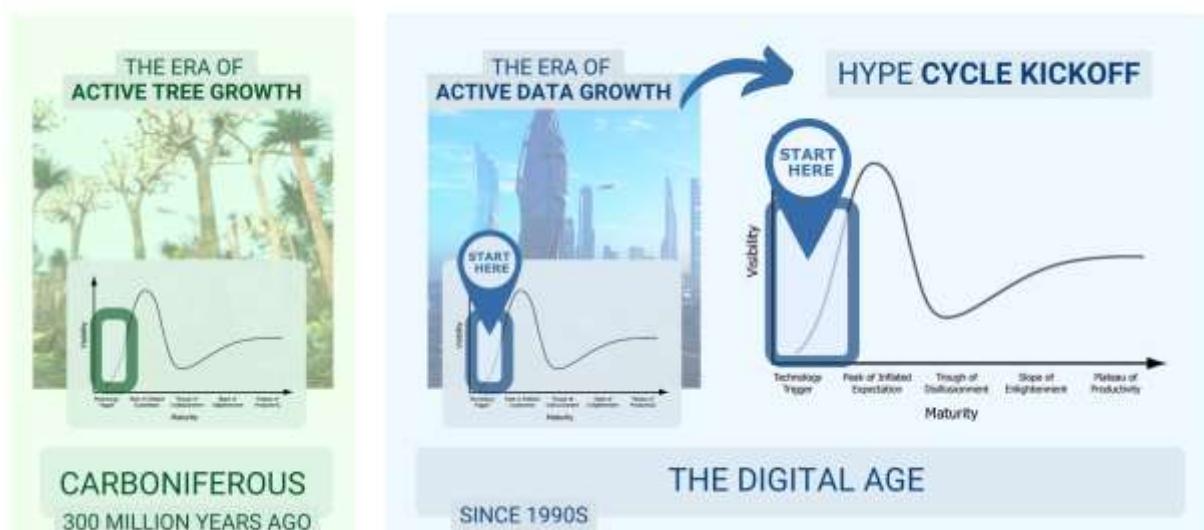
DIGITÁLNÍ REVOLUCE A EXPLOZE DAT

Počátek datového boomu jako evoluční vlna

Stavebnictví zažívá nebývalou informační explozi. Pokud si představíme podnikání jako strom znalostí (obr. 1.2-5), který je napájen daty, lze současnou fázi digitalizace přirovnat k rychlému růstu vegetace v období karbonu, tedy v éře, kdy se biosféra Země proměnila rychlým nahromaděním biomasy (obr. 1.3-1).

S globální digitalizací se množství informací ve stavebnictví každoročně zdvojnásobuje. Moderní technologie umožňují shromažďovat data na pozadí, analyzovat je v reálném čase a využívat je v rozsahu, který se ještě před nedávнем zdál nemožný.

Podle Moorova zákona, který formuloval Gordon Moore (spoluzakladatel společnosti Intel®), se hustota a složitost integrovaných obvodů a množství zpracovávaných a ukládaných dat zdvojnásobuje přibližně každé dva roky [19].



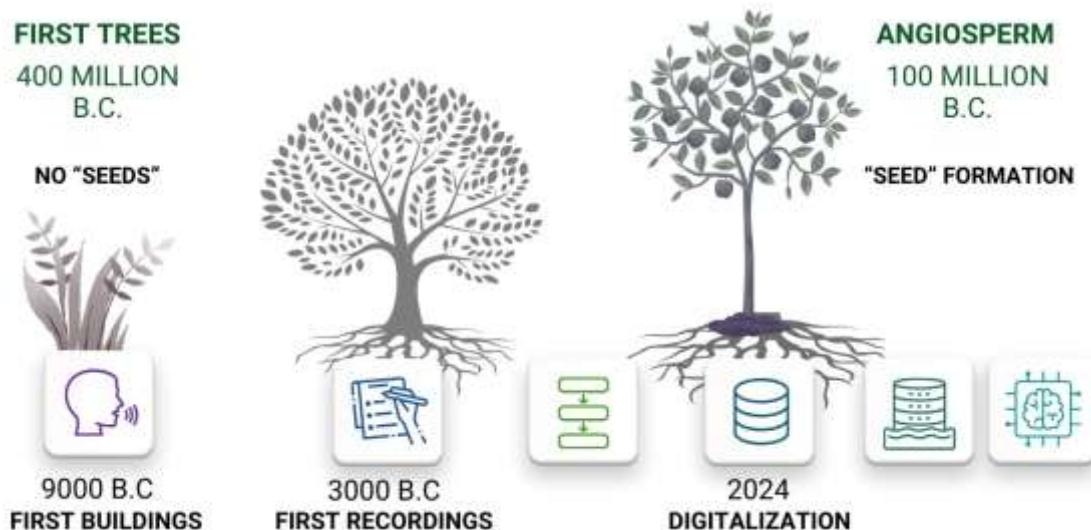
Obr. 1.3-1 Nástup digitalizace vedl k exponenciálnímu nárůstu dat, podobně jako prudký nárůst vegetace v době uhelné.

Zatímco starověké megalitické stavby, jako je Göbekli Tepe (Turecko), po sobě nezanechaly zdokumentované znalosti vhodné k opakování použití, dnešní digitální technologie umožňují shromažďovat a opakovat používat informace. Lze to přirovnat k evolučnímu přechodu od výtrusných rostlin k semenným rostlinám (nahosemenným): vznik semen dal podnět k rozsáhlému rozšíření života na planetě. (Obr. 1.3-2).

Podobně se data z minulých projektů stávají jakýmsi "digitálním semenem" - nositelem znalostí DNA, které lze rozšiřovat a používat v nových projektech a pro ducích. Nástup moderních nástrojů umělé inteligence - strojového učení a velkých jazykových modelů (LLM), jako jsou ChatGPT, LlaMa, Mistral,

Claude, DeepSeek, QWEN, Grok - umožňuje data automaticky extrahovat, interpretovat a používat v nových souvislostech

Stejně jako semena způsobila revoluci v šíření života na původně neživé planetě, stávají se "datová semena" základem pro automatický vznik nových informačních struktur a znalostí, což umožňuje digitálním ekosystémům vyvíjet se nezávisle a přizpůsobovat se měnícím se požadavkům uživatelů.



Obr. 1.3-2 Digitální "datová semena" hrají stejnou evoluční roli jako krytosemenné rostliny, které změnily ekosystém Země.

Stojíme na prahu nové éry ve stavebnictví, kde exploze dat a aktivní šíření "datových zárodků" - strukturovaných informací z minulých a probíhajících projektů - tvoří základ digitální budoucnosti odvětví. Jejich "opylování" prostřednictvím velkých datových jazykových modelů (LLM) nám umožňuje nejen pozorovat digitální změny, ale aktivně se podílet na vytváření samoučících se, adaptivních ekosystémů. To není evoluce - je to digitální revoluce, v níž se data stávají hlavním stavebním kamenem nové reality

Množství dat ve stavebnictví dramaticky roste díky informacím z různých oborů v průběhu celého životního cyklu stavebních projektů. Toto obrovské nahromadění dat posunulo stavebnictví do éry velkých dat [20].

- Prof. Hang Yang, katedra stavebního inženýrství a architektury, Wuhanská technologická univerzita, Wuhan.,

Růst objemu dat v informačním věku připomíná evoluční procesy v přírodě: stejně jako vývoj lesů změnil dávnou krajinu planety, současná informační exploze mění krajinu celého stavebnictví.

Množství dat generovaných v moderní společnosti

V posledních dvou letech bylo vytvořeno 90% všech existujících dat na světě [21]. Od roku 2023 každý člověk, včetně profesionálů ve stavebnictví, vygeneruje přibližně 1,7 megabajtu dat za sekundu [22] a celkový objem dat na světě dosáhne v roce 2023 64 zettabajtů a předpokládá se, že do roku 2025 překročí 180 zettabajtů, tedy $180 \cdot 10^{15}$ megabajtů [23].

Tato informační exploze má historický precedens - vynález knihtisku Johannesem Gutenbergem v 15. století. Pouhých padesát let po jeho zavedení se počet knih v Evropě zdvojnásobil: během několika desetiletí bylo vytisknuto kolik knih, kolik jich bylo vytvořeno ručně za předchozích 1 200 let [24]. Dnes jsme svědky ještě rychlejšího růstu: množství dat na světě se zdvojnásobuje každé tři roky.

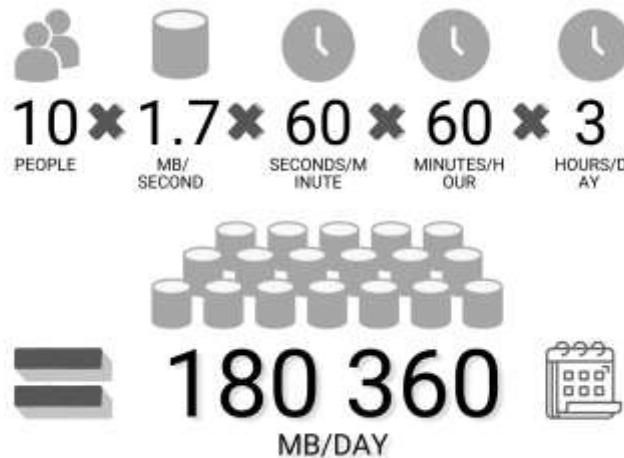
Vzhledem k současnemu tempu růstu objemu dat má stavebnictví potenciál vygenerovat během několika příštích desetiletí kolik informací, kolik jich nashromáždilo za celou svou předchozí historii



Obr. 1.3-3 Denní ukládání dat každým zaměstnancem na firemní servery přispívá k neustálému nárůstu dat.

V dnešním světě stavebnictví vytvářejí i malé firmy denně obrovské množství multiformátových informací a digitální stopa i malé stavební firmy může dosáhnout desítek gigabajtů denně - od modelů a výkresů až po fotografické záznamy a senzory na stavbě. Předpokládáme-li, že každý technik vygeneruje v průměru asi 1,7 MB dat za sekundu, odpovídá to přibližně 146 GB za den, resp. 53 TB za rok (obr. 1.3-3).

Pokud tým 10 lidí aktivně pracuje pouze 3 hodiny denně, kumulativní objem informací vytvořených za den dosahuje 180 gigabajtů (obr. 1.3-4).



Obrázek 1.3-4 Společnost s 10 zaměstnanci generuje přibližně 50-200 gigabajtů dat denně.

Za předpokladu, že 30% pracovních dat je nových (zbytek je přepsán nebo smazán), může firma o 10 lidech vytvořit řádově několik stovek gigabajtů nových dat měsíčně (skutečná čísla závisí na typu činnosti firmy)

Je tedy zřejmé, že nejenže generujeme stále více dat, ale také čelíme rostoucí potřebě jejich efektivní správy, ukládání a dlouhodobé dostupnosti. A zatímco dříve mohla data bezplatně "ležet" na lokálních serverech, v souvislosti s digitální transformací začíná stále více firem využívat cloudová řešení jako základ své informační infrastruktury.

Náklady na ukládání dat: ekonomické hledisko

V posledních letech stále více společností zadává ukládání dat do cloudových služeb. Pokud například společnost hostuje polovinu svých dat v cloudu při průměrné ceně 0,015 USD za gigabajt měsíčně, mohou se její náklady na ukládání dat každý měsíc zvýšit o 10-50 USD [25].

Pro malou společnost s typickými vzorci generování dat se mohou náklady na cloudové úložiště pohybovat v řádu stovek až potenciálně přes tisíc dolarů měsíčně (obrázek 1.3-5) v několika letech, což představuje potenciálně značnou finanční zátěž.

Podle studie společnosti Forrester "Enterprises Outsource Data Storage as Complexity Grows" [26], která se dotazovala 214 osob s rozhodovací pravomocí v oblasti technologické infrastruktury. [26], která se dotazovala 214 osob s rozhodovacími pravomocemi v oblasti technologické infrastruktury, více než třetina organizací outsourcuje datová úložiště, aby se vyrovnila s rostoucím objemem a složitostí datových operací, přičemž téměř dvě třetiny podniků dávají přednost modelu založenému na předplatném.



Obr. 1.3-5 Přesun dat do cloutu může zvýšit měsíční náklady na úložiště až o 2 000 dolarů, a to i u společnosti s pouhými 10 zaměstnanci.

Situaci dále komplikuje zrychlené zavádění cloudových technologií, jako jsou CAD (BIM), CAFM, PMIS a ERP -systémy, které dále zvyšují náklady na ukládání a zpracování dat. V důsledku toho jsou společnosti nuteny hledat způsoby, jak optimalizovat náklady a snížit závislost na poskytovatelích cloudových služeb.

Od roku 2023 se s aktivním vývojem velkých jazykových modelů (LLM) začaly měnit přístupy k ukládání dat. Stále více společností uvažuje o tom, že by nad svými daty převzaly zpět kontrolu, protože je bezpečnější a výhodnější zpracovávat informace na vlastních serverech.

V této souvislosti vystupuje do popředí trend odklonu od cloudového ukládání a zpracování pouze nezbytných dat ve prospěch lokálního nasazení podnikových LLM a AI -řešení. Jak jednom z rozhovorů [27] zdůraznil generální ředitel společnosti Microsoft, namísto spoléhání se na několik samostatných aplikací nebo cloudových řešení SaaS k provádění různých úkolů budou agenti AI řídit procesy v databázích a automatizovat funkce různých systémů.

[...] Starý přístup k této problematice [zpracování dat] byl takový, že když si vzpomenete, jak různé podnikové aplikace řešily integraci, používaly konektory. Společnosti prodávaly licence na tyto konektory a kolem toho se vytvořil obchodní model. SAP [ERP] je jedním z klasických příkladů: k datům SAP jste mohli přistupovat pouze tehdy, pokud jste měli správný konektor. Zdá se mi tedy, že něco podobného vznikne i v případě interakce agentů [AI] [...]. Přístup, alespoň ten, který zastáváme my, je následující: Myslím, že v éře agentů [AI] se pravděpodobně zhroutí koncept existence podnikových aplikací. Protože když se nad tím zamyslíte, jsou to v podstatě databáze s hromadou obchodní logiky

- Satya Nadella, generální ředitel společnosti Microsoft, rozhovor pro kanál BG2, 2024. [28]

V tomto paradigmatu jde přístup LLM založený na datech nad rámec klasických systémů. Umělá inteligence se stává prostředníkem mezi uživatelem a daty (obr. 2.2-3, obr. 2.2-4), čímž se eliminuje

potřeba více zprostředkujících rozhraní a zvyšuje se efektivita podnikových procesů. O tomto přístupu k práci s daty budeme více hovořit v kapitole "Proměna chaosu v řád a snižování složitosti".

Zatímco architektura budoucnosti se teprve utváří, společnosti se již nyní potýkají s důsledky minulých rozhodnutí. Masivní digitalizace posledních desetiletí doprovázená zaváděním různorodých systémů a nekontrolovaným hromaděním dat vedla k novému problému - informačnímu přetížení.

Hranice akumulace dat: od hmoty k významu

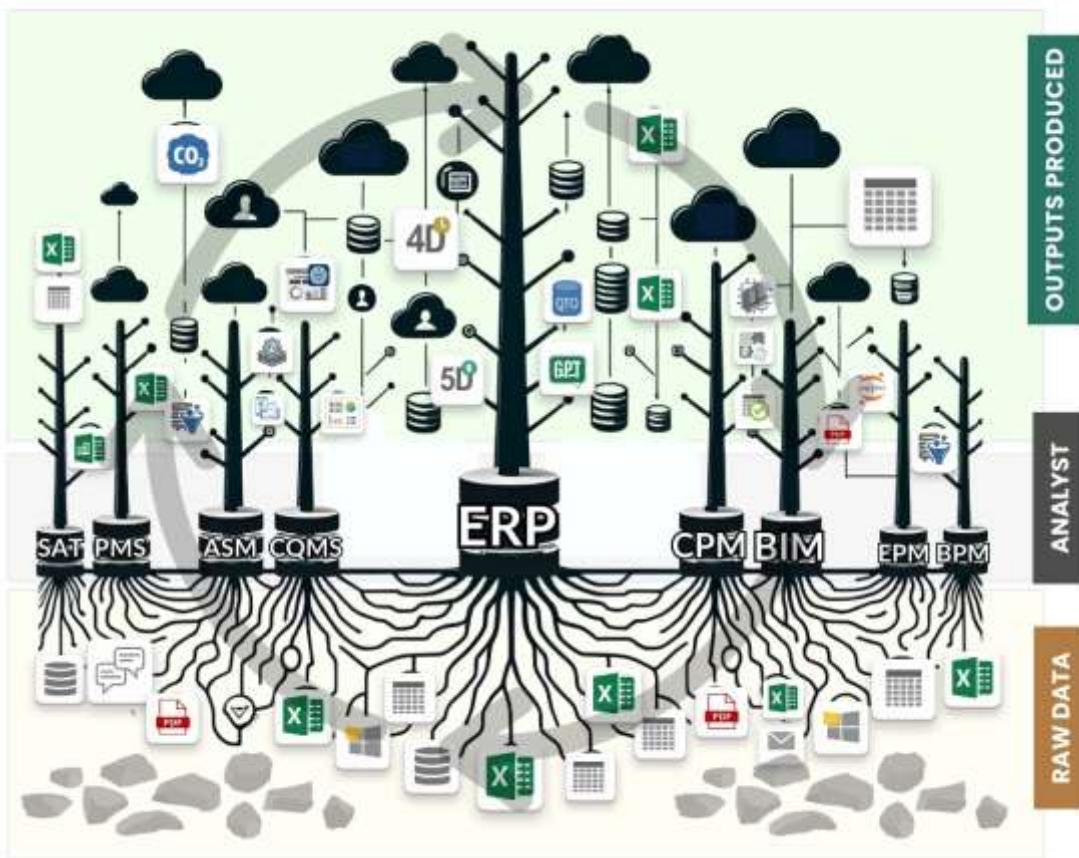
Moderní podnikové systémy se úspěšně rozvíjejí a fungují v podmínkách řízeného růstu, kdy je objem dat a počet aplikací v rovnováze s možnostmi IT oddělení a manažerů. V posledních desetiletích však digitalizace vedla k nekontrolovatelnému nárůstu objemu a složitosti dat, což způsobilo efekt přesycení informačního ekosystému firem.

Servery a úložiště jsou dnes vystaveny nebývalému přílivu nezpracovaných a různě formátovaných informací, které se nestihnou proměnit v kompost a rychle ztrácejí význam. Omezené podnikové zdroje si s touto záplavou nedokážou poradit a data se hromadí v izolovaných silech (tzv. "silos"), která vyžadují ruční zpracování, aby z nich bylo možné získat užitečné informace.

Moderní systémy řízení firem tak často trpí informačním přetížením, podobně jako les porostlý břečtanem a pokrytý plísní. Místo výživného informačního humusu v srdci podnikového ekosystému vznikají izolované oblasti dat různého formátu, což nevyhnutelně vede ke snížení celkové efektivity podnikových procesů.

Po dlouhém období exponenciálního růstu dat v posledních 40 letech bude nevyhnutelně následovat fáze nasycení a následného ochlazení. Až úložiště dosáhne svých limitů, dojde ke kvalitativnímu posunu: data již nebudou pouhým objektem pro ukládání, ale strategickým zdrojem.

S rozvojem umělé inteligence a strojového učení mají společnosti možnost snížit náklady na zpracování informací a přejít od kvantitativního růstu ke kvalitativnímu využití dat. V příštím desetiletí se bude muset stavebnictví přesunout od vytváření stále většího množství dat k zajištění jejich struktury, integrity a analytické hodnoty.



Obrázek 1.3-6 Izolované zdroje dat brání sdílení informací mezi datovými systémy.

Hlavní hodnota již nespočívá v množství informací, ale ve schopnosti je automaticky interpretovat a přeměnit na aplikované znalosti užitečné pro manažerská rozhodnutí. Aby se data stala skutečně užitečnými, musí být správně spravována: shromažďována, ověřována, strukturována, ukládána a analyzována v kontextu konkrétních obchodních úkolů.

Proces analýzy dat ve společnosti se podobá cyklu života a rozpadu stromů v lese a vzniku nových mladých a silných stromů: dospělé stromy odumírají, rozkládají se a stávají se živou půdou pro nový růst. Hotové a dokončené procesy se po svém dokončení stávají součástí informačního ekosystému společnosti a nakonec se stávají informačním humusem, který pohání budoucí růst nových systémů a dat.

V praxi je však tento cyklus často přerušen. Místo organické obnovy vzniká vrstevnatý chaos podobný geologickým vrstvám, kdy se nové systémy vrství na staré bez hlubší integrace a strukturace. V důsledku toho vznikají nesourodá informační "sila", která brání oběhu znalostí a komplikují správu dat.

Další kroky: od teorie dat k praktickým změnám

Vývoj dat ve stavebnictví je cestou od hliněných tabulek k moderním modulárním platformám. Dnešní výzvou není shromažďování informací, ale vytvoření rámce, který z nesourodých a různorodých dat vytvoří strategický zdroj. Ať už je vaše role manažera společnosti nebo inženýra, pochopení hodnoty dat a práce s nimi bude v budoucnu klíčovou dovedností.

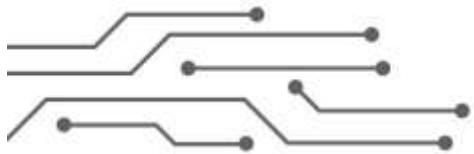
Pro shrnutí této části je vhodné zdůraznit hlavní praktické kroky, které vám pomohou aplikovat diskutované přístupy ve vašich každodenních úkolech:

- Proveďte osobní audit informačních toků
 - Vytvořte si seznam všech systémů a aplikací, se kterými denně pracujete.
 - Označte, kde trávíte nejvíce času vyhledáváním nebo překontrolováním dat.
 - Určete své klíčové zdroje informací
 - Analýza stávajícího prostředí aplikací z hlediska nadbytečnosti a duplicit funkcí.
- Snažte se procházet procesy podle úrovni analytické vyspělosti.
 - své úkoly popisnou analýzou (co se stalo?).
 - Postupně zavádějte diagnostiku (proč se to stalo?)
 - Přemýšlejte o tom, jak v procesech přejít k prediktivní (co se stane?) a preskriptivní (co dělat?) analytice.
- Začněte strukturovat pracovní data
 - Zavedení jednotného systému pro pojmenování souborů a složek, které často používáte při své práci.
 - Vytváření šablon pro často používané dokumenty a zprávy
 - Pravidelně archivujte dokončené projekty s jasnou strukturou.

I když nemůžete změnit celou informační infrastrukturu ve svém týmu nebo společnosti, začněte s vlastními procesy a drobnými zlepšeními v každodenní práci. Pamatujte, že skutečná hodnota dat nespočívá v jejich objemu, ale ve schopnosti získat z nich užitečné poznatky. I malé, ale správně strukturované a analyzované soubory dat mohou mít významný dopad, pokud jsou začleněny do rozhodovacích procesů.

V dalších částech knihy přejdeme ke konkrétním metodám a nástrojům pro práci s daty, podíváme se na způsoby transformace nestrukturovaných informací do strukturovaných souborů, prozkoumáme technologie pro automatizaci analytiky a podrobně popíšeme, jak vybudovat efektivní analytický ekosystém ve stavební společnosti.





II ČÁST

JAK SE STAVEBNICTVÍ UTÁPÍ V DATOVÉM CHAOSU

Druhá část je věnována kritické analýze výzev, kterým čelí stavební společnosti při práci s rostoucím objemem dat. Podrobně jsou zkoumány důsledky roztríštěnosti informací a fenomén "dat v silech", který brání efektivnímu rozhodování. Zkoumá se problematika HiPPO -přístupu (Highest Paid Person's Opinion) a jeho dopad na kvalitu manažerských rozhodnutí ve stavebních projektech. Posuzuje se vliv dynamických podnikových procesů a jejich rostoucí složitosti na informační toky a provozní efektivitu. Jsou uvedeny konkrétní příklady toho, jak nadměrná složitost systémů zvyšuje náklady a snižuje flexibilitu organizací. Zvláštní pozornost je věnována omezením způsobeným proprietárními formáty a perspektivám využití otevřených standardů ve stavebnictví. Je představena koncepce přechodu k softwarovým ekosystémům založeným na umělé inteligenci a LLM, které minimalizují nadměrnou složitost a technické překážky.

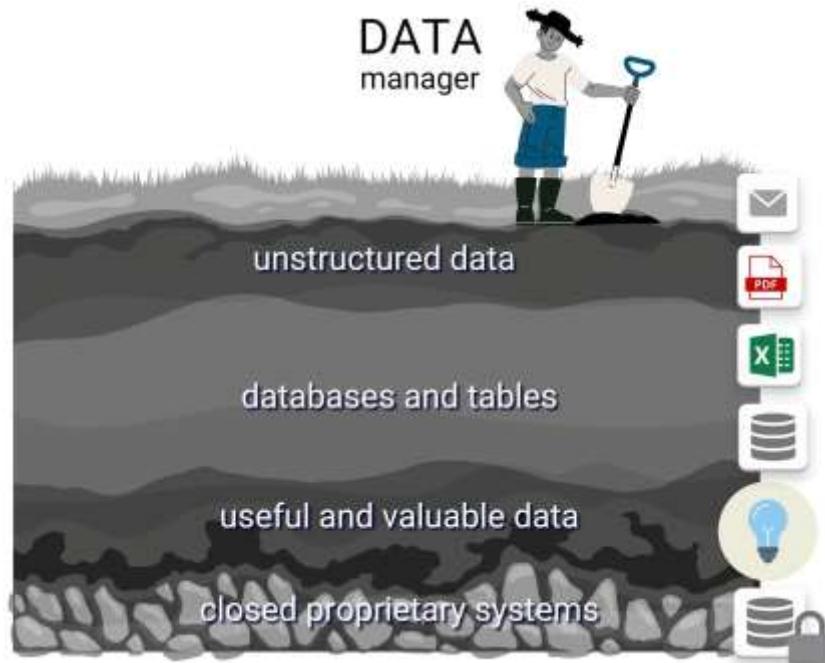
KAPITOLA 2.1.

ROZTŘÍSTĚNOST DAT A DATOVÁ SILA

Čím více nástrojů, tím efektivnější podnikání?

Na první pohled se může zdát, že více digitálních nástrojů vede k vyšší efektivitě. V praxi tomu tak ale není. S každým novým řešením, ať už se jedná o cloudovou službu, starší systém nebo další sestavu v Excelu, přidává společnost do svého digitálního prostředí další vrstvu - vrstvu, která často není integrována s ostatními (obrázek 2.1-1).

Data lze přirovnat k uhlí nebo ropě: hromadí se léta, zhutňují se pod vrstvami chaosu, chyb, nestrukturovaných procesů a zapomenutých formátů. Aby z nich firmy získaly skutečně užitečné informace, musí se doslova prodírat vrstvami zastaralých řešení a digitálního šumu.



Obr. 2.1-1 Různorodá data tvoří rozčleněné vrstvy - i "zlaté" poznatky se ztrácejí v geologických horninách systémové složitosti.

Každá nová aplikace za sebou zanechává stopu: soubor, tabulkou nebo celé izolované "silo" na serveru. Jedna vrstva je hlína (zastaralá a zapomenutá data), druhá písek (nesourodé tabulky a sestavy) a třetí žula (uzavřené proprietární formáty, které nelze integrovat). Postupem času se digitální prostředí společnosti stále více podobá zásobníku nekontrolovaného hromadění informací, kde se hodnota ztrácí hluboko v podnikových serverech.

S každým novým projektem a každým novým systémem se stává složitější nejen infrastruktura, ale také cesta k užitečným kvalitním datům. Dostat se k cennému "kameni" vyžaduje hloubkové čištění,

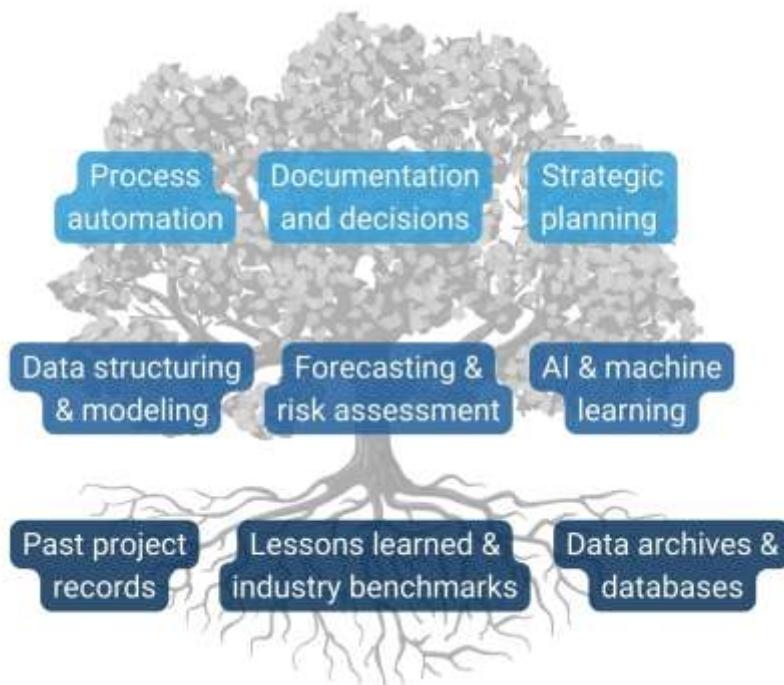
strukturování informací, "chunking", jejich seskupování do smysluplných celků a extrakci strategicky důležitých poznatků prostřednictvím analytiky a datového modelování.

Data jsou cenná věc a vydrží déle než samotné systémy [které data zpracovávají] [29].

- Tim Berners-Lee, otec World Wide Webu a tvůrce první webové stránky

Než se data stanou "cennou věcí" a spolehlivým podkladem pro rozhodování, musí projít pečlivou přípravou. Právě správné předzpracování promění různorodá data ve strukturovanou zkušenosť, užitečný informační humus, který se pak stane nástrojem pro prognózování a optimalizaci.

Panuje mylná představa, že k zahájení analýzy potřebujete dokonale čistá data, ale v praxi je schopnost pracovat se špinavými daty nezbytnou součástí procesu.



Obr. 2.1-2 Data jsou kořenovým systémem a základem podnikání, které zase vychází z rozhodovacích procesů.

S tím, jak se technologie neustále vyvíjejí, musí i vaše firma postupovat vpřed a naučit se vytvářet hodnotu z dat. Stejně jako ropné a uhelné společnosti budují infrastrukturu pro těžbu nerostných surovin, musí se i podniky naučit, jak řídit tok nových informací na svých vlastních serverech a získávat cenné poznatky z nepoužívaných, neformátovaných a zastaralých dat a přeměnit je ve strategický zdroj.

Prvním krokem je vytvoření polí (datových skladů). Ani ty nejvýkonnější nástroje nevyřeší problém izolace dat a víceformátových dat, pokud budou společnosti nadále pracovat v oddělených systémech. Pokud data existují odděleně od sebe, aniž by se vzájemně prolínala a sdílela informace, celí podniky efektu "datového sila". Namísto jednotné a konzistentní infrastruktury jsou společnosti nuceny vynakládat prostředky na slučování a synchronizaci dat.

Datová sila a jejich dopad na výkonnost společnosti

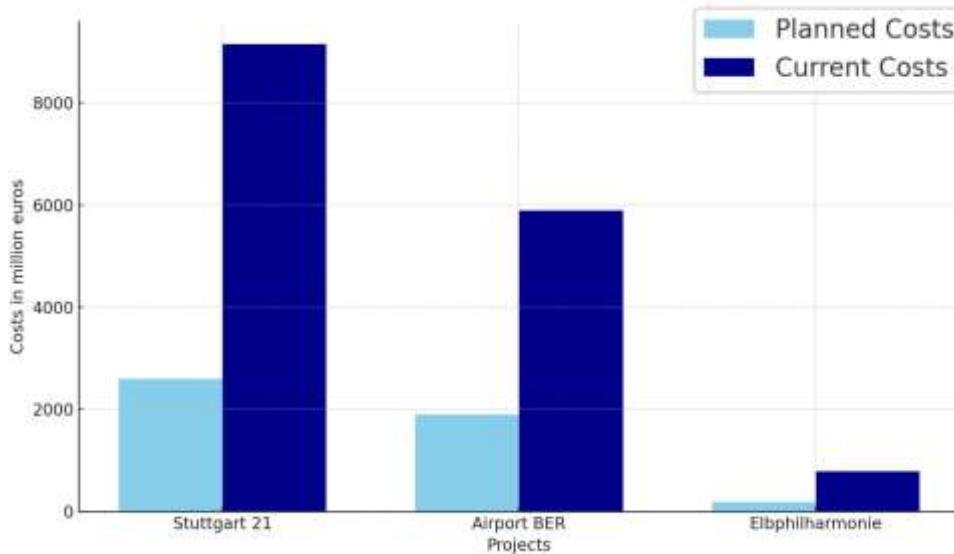
Představte si, že stavíte sídliště, ale každý tým má svůj vlastní projekt. Některí staví zdi, jiní pokládají komunikace a další dláždí silnice, aniž by se vzájemně kontrolovali. Výsledkem je, že potrubí neodpovídá otvorům ve zdech, výtahové šachty neodpovídají podlažím a silnice se musí rozebrat a položit znova.

Tato situace není jen hypotetickým scénářem, ale realitou mnoha moderních stavebních projektů. Vzhledem k velkému počtu generálních a subdodavatelských firem pracujících s různými systémy a bez jednotného koordinačního centra se proces mění v sérii nekonečných schvalování, přepracování a konfliktů. To vše vede ke značným zpožděním a několikanásobným nákladům na projekt.

Klasická situace na staveništi je jednoduchá: bednění je hotové, ale dodávka výztuže nedorazila včas. Při kontrole informací v různých systémech probíhá komunikace zhruba takto:

- ⌚ **Mistr** na stavbě 20. dne píše vedoucímu projektu: "*Dokončili jsme bednění, kde je armatura?*".
- ⌚ **Projektový manažer** (PMIS) oddělení veřejných zakázek: - "*Bednění je připraveno. V mé systému [PMIS] se píše, že výztuž měla dorazit 18. září. Kde jsou výztuže?*"
- ⌚ **Specialista dodavatelského řetězce** (ERP): - "*Náš ERP říká, že dodávka bude 25.*"
- ⌚ **Datový inženýr** nebo IT oddělení (zodpovědné za integrace): - V PMIS je datum 18., v ERP je to 25. Mezi ERP a PMIS neexistuje propojení OrderID, takže data nejsou synchronizována. Toto je typický příklad informační mezery.
- ⌚ **Projektový manažer** generálnímu řediteli - "*Dodávka kování se zpožďuje, staveniště stojí a není jasné, kdo je za to zodpovědný*".

Příčinou incidentu byla izolace dat v různých systémech. Integrací a sjednocením datových zdrojů, vytvořením jediného úložiště informací a automatizací prostřednictvím nástrojů ETL (Apache NiFi, Airflow nebo n8n) lze odstranit silážní místa mezi systémy. Tyto a další metody a nástroje budou podrobně rozebrány v dalších částech knihy.



Obr. 2.1-3 Srovnání plánovaných a skutečných nákladů na velké infrastrukturní projekty v Německu.

Totéž se děje s podnikovými systémy: nejprve se vytvoří izolovaná řešení a pak se musí vynaložit obrovské prostředky na jejich integraci a harmonizaci. Kdyby byly datové a komunikační modely promyšleny od začátku, nebylo by integrace vůbec zapotřebí. Oddělená data vytvářejí v digitálním světě chaos, podobně jako nekoordinovaný stavební proces.

Podle studie KPMG 2023 "Cue construction 4.0: Time to make or break" pouze 36% společností efektivně sdílí data napříč odděleními, zatímco 61% se potýká s vážnými problémy kvůli izolovaným datovým "silům" [30].



Obr. 2.1-4 Léta obtížně shromažďovaná data se hromadí v izolovaných úložných "silech" s rizikem, že nebudou nikdy využita.

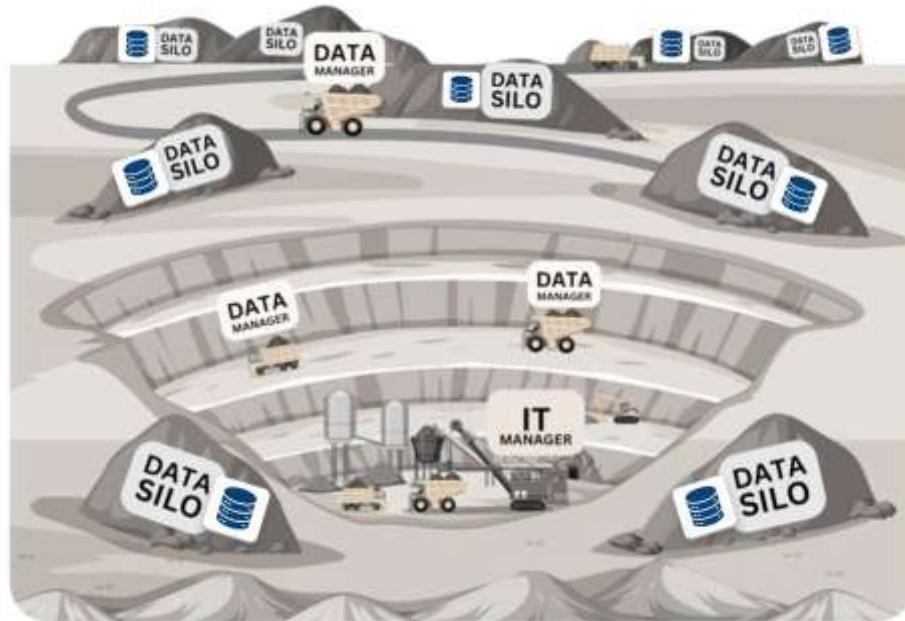
Firemní data jsou uložena v izolovaných systémech jako jednotlivé stromy roztroušené po krajině. Každý z nich obsahuje cenné informace, ale nedostatek vazeb mezi nimi brání vytvoření jednotného, vzájemně propojeného ekosystému. Tato oddělenost brání toku dat a omezuje schopnost organizace vidět celý obraz. Propojení těchto sil je nesmírně dlouhý a složitý proces pěstování podhoubí na úrovni managementu, aby se naučil přenášet jednotlivé informace mezi systémy.

Podle studie WEF z roku 2016 je jednou z hlavních překážek digitální transformace nedostatek společných datových standardů a roztríštěnost.

Stavebnictví je jedním z nejvíce roztríštěných odvětví na světě a závisí na hladké interakci všech účastníků hodnotového řetězce [5].

- Světové ekonomické fórum 2016: Formování budoucnosti

Návrháři, manažeři, koordinátoři a vývojáři často dávají přednost samostatné práci, aby se vyhnuli složité koordinaci. Tento přirozený sklon vede k vytváření informačních "sil", v nichž jsou data izolována v rámci samostatných systémů. Čím více takových izolovaných systémů existuje, tím obtížnější je přimět je ke spolupráci. Časem každý systém získá vlastní databázi a specializované podpůrné oddělení manažerů (obrázek 1.2-4), což integraci dále komplikuje.



Obr. 2.1-5 Každý systém má tendenci vytvářet vlastní unikátní silu dat, která je třeba zpracovávat vhodnými nástroji [31].

Začarovaný kruh v podnikových systémech vypadá následovně: firmy investují do složitých iso leated řešení, pak čelí vysokým nákladům na jejich integraci a vývojáři, kteří si uvědomují složitost kombinování systémů, raději pracují ve svých uzavřených ekosystémech. To vše zvyšuje roztríštěnost IT prostředí a ztěžuje přechod na nová řešení (obrázek 2.1-5). Manažeři nakonec kritizují datová sila,

ale málokdy analyzují jejich příčiny a způsoby, jak jim předcházet. Manažeři si stěžují na zastaralé IT systémy, ale jejich nahrazení vyžaduje značné investice a málokdy přináší očekávané výsledky. V důsledku toho i pokusy o boj s problémem často situaci ještě zhoršují.

Hlavním důvodem odpojení je upřednostňování aplikací před daty. Společnosti nejprve vyvýjejí samostatné systémy nebo nakupují hotová řešení od dodavatelů a poté se je snaží sjednotit vytvořením duplicitních a nekompatibilních úložišť a databází.

Překonání problému fragmentace vyžaduje radikální nový přístup - upřednostnění dat před aplikacemi. Společnosti musí nejprve vypracovat strategie správy dat a datové modely a poté vybudovat systémy nebo zakoupit řešení, která budou pracovat s jediným souborem informací, a ne vytvářet nové překážky.

Vstupujeme do nového světa, kde jsou data možná důležitější než software.

- Tim O'Reilly, generální ředitel společnosti O'Reilly Media, Inc.

Studie McKinsey Global Institute "Rethinking Construction: the path to improved productivity" (2016) ukazuje, že stavebnictví zaostává za ostatními odvětvími v digitální transformaci [32]. Podle této zprávy může zavedení automatizované správy dat a digitálních platform výrazně zvýšit produktivitu a snížit ztráty spojené s nekonzistencí procesů. Tuto potřebu digitální transformace zdůrazňuje také zpráva Egan (Velká Británie, 1998) [33], která zdůrazňuje klíčovou roli integrovaných procesů a kolaborativního přístupu ve stavebnictví.

Výsledkem je, že zatímco v posledních 10 000 letech byl hlavním problémem správců dat nedostatek dat, s lavinou dat a systémů pro správu dat se uživatelé a správci potýkají s problémem - přemírou dat, což ztěžuje nalezení právně správných a kvalitních informací.

Rozdílná datová sila nevyhnutelně vedou k závažnému problému snížené kvality dat. V případě více nezávislých systémů mohou tytéž údaje existovat v různých verzích, často s protichůdnými hodnotami, což vytváří další potíže pro uživatele, kteří musí určit, které informace jsou relevantní a spolehlivé.

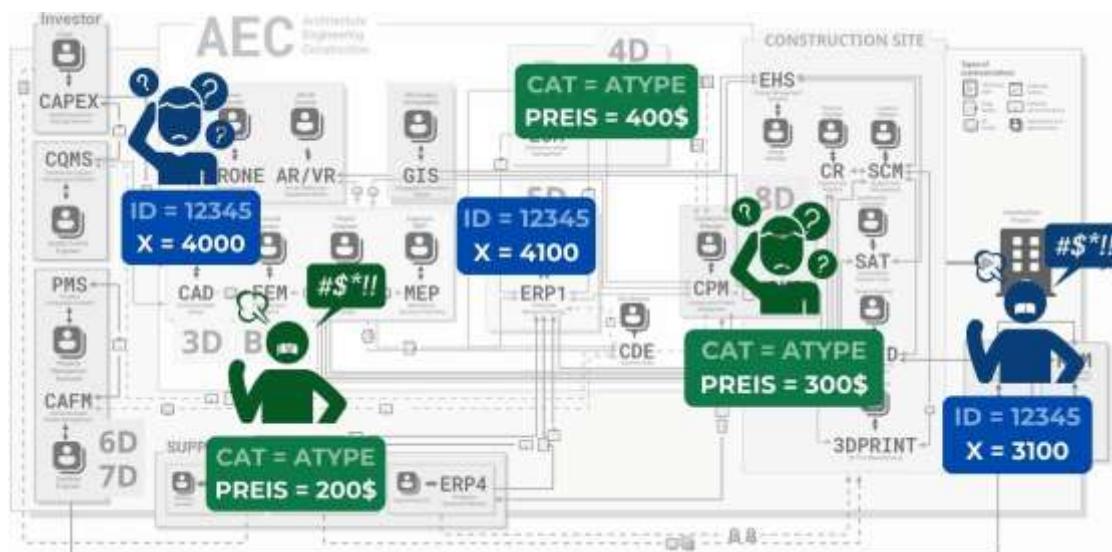
Duplicita a nedostatečná kvalita dat jako důsledek nejednotnosti.

Kvůli problému datových sil musí manažeři trávit značný čas vyhledáváním a porovnáváním dat. Aby se společnosti pojistily proti problémům s kvalitou, vytvářejí složité struktury řízení informací, v nichž je za vyhledávání, ověřování a odsouhlasování dat odpovědná vertikála manažerů. Tento přístup však pouze zvyšuje byrokracii a zpomaluje rozhodování. Čím více je dat, tím obtížnější je jejich analýza a interpretace, zejména pokud neexistuje jednotný standard pro jejich ukládání a zpracování.

S množstvím softwarových aplikací a systémů, které v posledním desetiletí rostou jako houby po dešti, je pro koncové uživatele stále důležitější problém sil a nevhodné kvality dat. Stejná data, ale s různými

hodnotami, lze nyní nalézt v různých systémech a aplikacích (obrázek 2.1-6). To vede k potížím pro koncové uživatele, když se snaží určit, která verze dat je z mnoha dostupných relevantní a správná. To vede k chybám v analýzách a nakonec i v rozhodování.

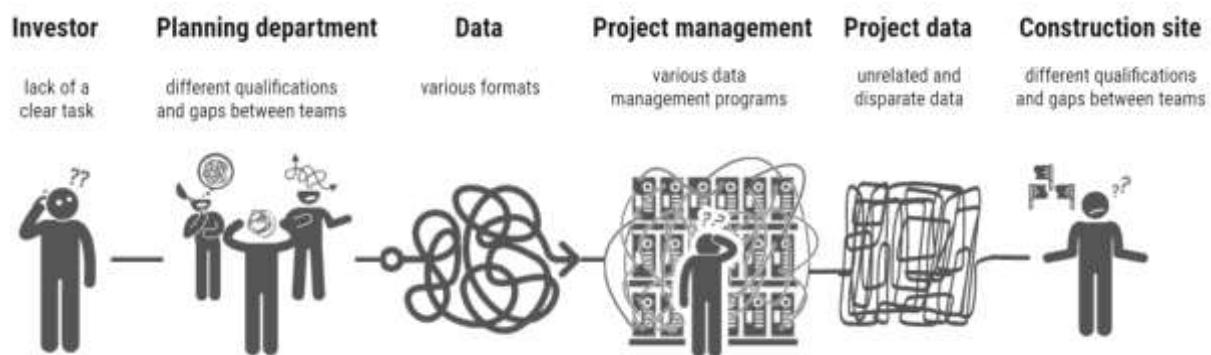
Aby se manažeři společnosti pojistili proti problémům s vyhledáváním správných údajů, vytvářejí víceúrovňovou byrokracií ověřovacích manažerů. Jejich úkolem je umět rychle vyhledat, zkontrolovat a odeslat požadovaná data ve formě tabulek a výkazů, přičemž se musí orientovat ve spletí různorodých systémů.



Obr. 2.1-6 Při snaze najít správná data musí manažeři zajistit kvalitu a právní spolehlivost dat mezi různými systémy.

V praxi však tento model přináší nové složitosti. Pokud jsou data spravována ručně a informace jsou rozptýleny v mnoha nesouvisejících rozhodnutích, stává se každý pokus o získání přesných a aktuálních informací prostřednictvím pyramidy rozhodovacích orgánů (obrázek 2.1-7) úzkým místem - časově náročným a náchylným k chybám.

Situaci zhoršuje lavina digitálních řešení. Softwarový trh je stále zaplaven novými nástroji, které se zdají být slibné. Ale bez jasné strategie správy dat se tato řešení neintegrují do jednotného systému, ale naopak vytvářejí další vrstvy složitosti a duplicity. Výsledkem je, že namísto zjednodušení procesů se firmy ocitají v ještě roztríštěnějším a chaotičtějším informačním prostředí.



Obr. 2.1-7 Složitost systémů a různorodost datových formátů vede ke ztrátě konzistence v procesu konstrukce.

Všechny tyto problémy spojené se správou množství různorodých řešení dříve či později přivedou vedení společnosti k důležitému zjištění: nejde o objem dat nebo hledání dalšího "univerzálního" nástroje pro zpracování dat. Skutečný důvod spočívá v kvalitě dat a ve způsobu, jakým je organizace vytváří, přijímá, ukládá a využívá.

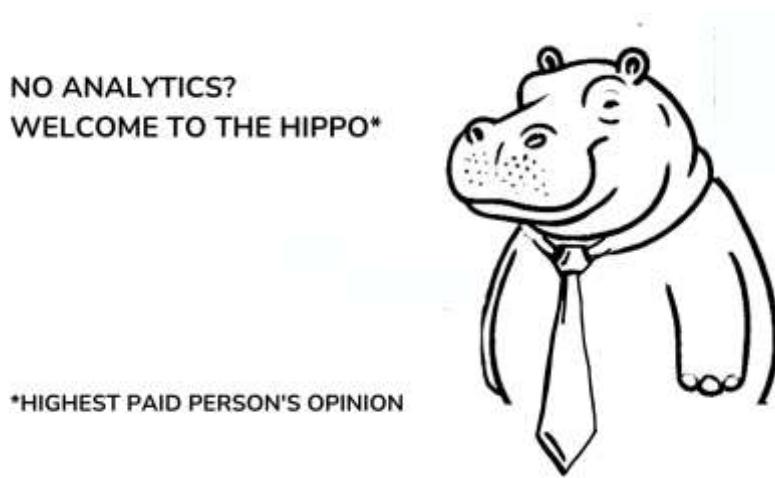
Klíčem k trvalému úspěchu není honba za novými "zázračnými" aplikacemi, ale budování datové kultury ve firmě. To znamená považovat data za strategické aktivum a učinit kvalitu, integritu a relevanci dat prioritou na všech úrovních organizace.

Řešení dilematu kvalita versus kvantita spočívá ve vytvoření jednotné datové struktury, která eliminuje duplicity, odstraňuje nekonzistence a sjednocuje informační toky. Tato architektura poskytuje jediný spolehlivý zdroj dat, na jehož základě lze činit informovaná, přesná a včasná rozhodnutí.

V opačném případě, jak se stále často stává, se společnosti nadále spoléhají spíše na subjektivní názory a intuitivní hodnocení odborníků na HiPPO než na spolehlivá fakta. Ve stavebnictví, kde odbornost tradičně hraje významnou roli, je to obzvláště patrné.

HiPPO aneb nebezpečí názorů při rozhodování

Ve stavebnictví se klíčová rozhodnutí tradičně přijímají na základě zkušeností a subjektivního úsudku. Bez včasných a spolehlivých údajů musí manažeři firem jednat naslepo a spoléhat se spíše na intuici nejlépe placených zaměstnanců (HiPPO - Highest Paid Person's Opinion) než na objektivní fakta (obr. 2.1-8).



Obrázek 2.1-8 Při absenci analytiky podnikání závisí na subjektivním názoru zkušených odborníků.

Tento přístup může být oprávněný ve stabilním a pomalu se měnícím prostředí, ale v éře digitální transformace se stává vážným rizikem. Rozhodnutí založená na intuici a odhadech jsou náchylná ke zkreslení, často vycházejí z nepodložených hypotéz a neberou v úvahu komplexní obraz odrázející se v datech

To, co je na úrovni rozhodování ve firmě vydáváno za inteligentní debatu, se často nezakládá na ničem konkrétním. Úspěch firmy by neměl záviset na autoritě a platu odborníků, ale na schopnosti efektivně pracovat s daty, identifikovat zákonitosti a činit informovaná rozhodnutí.

Je důležité opustit představu, že autorita nebo zkušenosť automaticky znamená, že rozhodnutí je správné. Přístup založený na datech mění pravidla hry: základem pro rozhodování jsou nyní data a analytika, nikoliv pozice a plat. Velká data, strojové učení a vizuální analytika nám umožňují identifikovat vzorce a spoléhat se na fakta, nikoli na dohad (obr. 1.1-4).

Bez dat jste jen další člověk s názorem [34].

- W. Edwards Deming, vědec a poradce v oblasti řízení

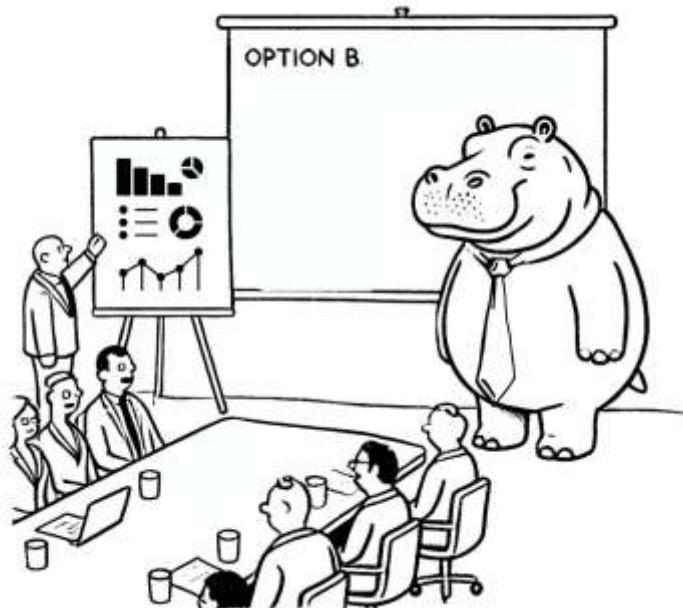
Moderní metody správy dat také zajišťují kontinuitu znalostí v rámci společnosti. Jasně popsané procesy, automatizace a systematický přístup umožňují předávat i klíčové role bez ztráty efektivity.

Slepá důvěra v data však může vést i k závažným chybám. Data sama o sobě jsou jen souborem čísel. Bez správné analýzy, kontextu a schopnosti identifikovat vzorce nemají žádnou hodnotu a nemohou řídit procesy. Klíč k úspěchu nespočívá ve volbě mezi intuicí HiPPO a analýzou, ale v budování inteligentních nástrojů, které přemění různorodé informace na zvládnutelná a informovaná rozhodnutí.

V digitálním stavebním prostředí není rozhodujícím faktorem úspěchu služební postup a místo v hierarchii, ale rychlosť reakce, přesnost rozhodování a efektivita zdrojů

Data jsou nástrojem, nikoli absolutní pravdou. Měla by doplňovat lidské myšlení, ne ho nahrazovat. Navzdory výhodám analytiky, data nemohou zcela nahradit lidskou intuici a zkušenosti. Jejich úloha spočívá v tom, že pomáhají přijímat přesnější a informovanější rozhodnutí.

Konkurenční výhody se dosáhne nejen dodržováním standardů, ale i tím, že dokážeme překonat konkurenci v efektivním využívání zdrojů, které jsou pro všechny stejné. V budoucnu budou datové dovednosti stejně důležité, jako kdysi byla gramotnost nebo matematické dovednosti. Odborníci, kteří budou umět analyzovat a interpretovat data, budou schopni činit přesnější rozhodnutí a vytlačit, kteří se spoléhají pouze na osobní zkušenosti (obrázek 2.1-9).



Obr. 2.1-9 Rozhodnutí by měla být založena na objektivní analýze, nikoli na názoru nejlépe placeného zaměstnance.

Manažeři, specialisté a inženýři budou působit jako datoví analytici, kteří budou zkoumat strukturu, dynamiku a klíčové ukazatele projektů. Lidské zdroje se stanou prvky systému, které budou vyžadovat flexibilní přizpůsobení na základě dat, aby se maximalizovala efektivita.

Chyby při použití nedostatečných dat jsou mnohem menší než při použití žádných dat [35].

- Charles Babbage, vynálezce prvního analytického počítacího stroje

Vznik velkých dat a zavedení LLM (Large Language Models) radikálně změnily nejen způsob analýzy, ale i samotnou podstatu rozhodování. Zatímco dříve se pozornost soustředila na kauzalitu (proč se něco stalo - diagnostická analytika) (obr. 1.1-4), dnes se do popředí dostává schopnost předvídat budoucnost (prediktivní analytika) a v budoucnu i preskriptivní analytika, kdy strojové učení a umělá inteligence navrhují nejlepší volbu v rozhodovacím procesu.

Podle nové studie společnosti SAP™ "New Study Finds Nearly Half of Executives Trust Artificial Intelligence More Than Themselves" z roku 2025 [36] by 44% vrcholových manažerů bylo ochotno změnit své předchozí rozhodnutí na základě rady umělé inteligence, a 38% by umělé inteligenci důvěřovalo, aby za ně činila obchodní rozhodnutí. Přitom 74% vedoucích pracovníků uvedlo, že důvěřují radám AI více než svým přátelům a rodině, a 55% pracuje ve společnostech, kde poznatky získané pomocí AI nahrazují nebo často obcházejí tradiční metody rozhodování - zejména v organizacích s ročními příjmy nad 5 miliard dolarů. Kromě toho 48% respondentů používá generativní nástroje AI denně, z toho 15% je používá několikrát denně.

S rozvojem LLM a automatizovaných systémů správy dat vyvstává nová výzva: jak efektivně využívat informace, aniž by ztratily svou hodnotu v chaosu nekompatibilních formátů a heterogenních zdrojů, který je doplněn rostoucí složitostí a dynamikou podnikových procesů.

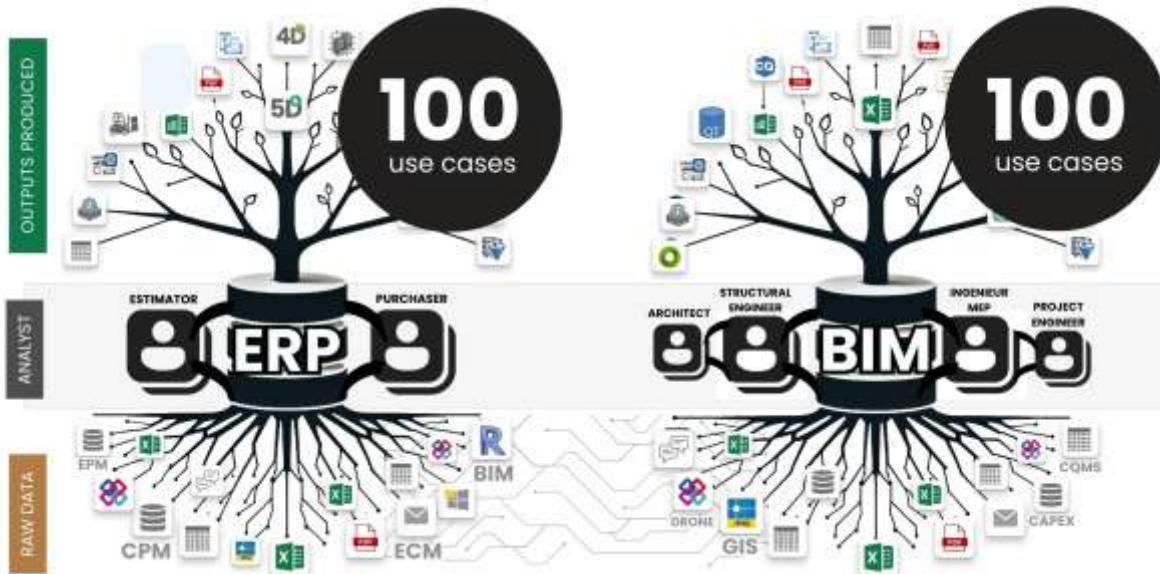
Neustálý nárůst složitosti a dynamiky podnikových procesů

Stavebnictví dnes čelí vážným výzvám v oblasti správy dat a procesů. Hlavními problémy jsou oddělené informační systémy, nadměrná byrokracie a nedostatečná integrace digitálních nástrojů. Tyto výzvy se stupňují s tím, jak se samotné podnikové procesy stávají složitějšími - pod vlivem technologií, měnících se požadavků zákazníků a vyvíjejících se předpisů.

Jedinečnost stavebních projektů je dána nejen jejich technickými zvláštnostmi, ale také rozdíly v národních normách a regulačních požadavcích v různých zemích (obr. 4.2-10, obr. 5.1-7). To vyžaduje flexibilní a individuální přístup ke každému projektu, který je v rámci tradičních modulárních řídicích systémů obtížně realizovatelný. Vzhledem ke složitosti procesů a velkému množství dat se mnoho společností obrací na dodavatele nabízející specializovaná řešení. Trh je však přetížený - mnoho začínajících firem nabízí podobné produkty, zaměřené na úzké úlohy. V důsledku toho se často vytrácí holistický přístup ke správě dat.

Rozhodujícím faktorem konkurenceschopnosti se stává přizpůsobení se neustálému přílivu nových technologií a požadavkům trhu. Stávající proprietární aplikace a modulární systémy však mají nízkou přizpůsobivost - jakékoli změny vyžadují často zdlouhavé a nákladné revize ze strany vývojářů, kteří ne vždy rozumí specifikům stavebních procesů.

Společnosti se stávají rukojmím technologického zpoždění a čekají na nové aktualizace, místo aby urychleně zaváděly inovativní integrované přístupy. V důsledku toho je vnitřní struktura stavebních organizací často složitým ekosystémem vzájemně propojených hierarchických a často uzavřených systémů koordinovaných prostřednictvím víceúrovňové sítě manažerů (obrázek 2.1-10).

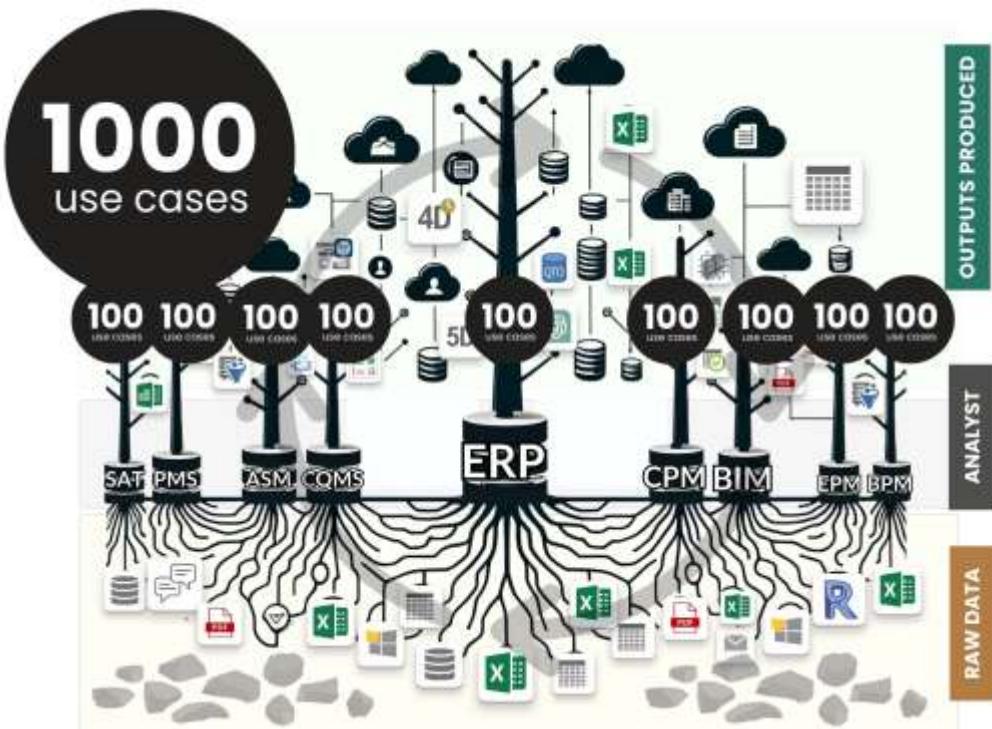


Obr. 2.1-10 Podniky jsou tvořeny vzájemně propojenými systémy, jejichž propojení tvoří procesy, které vyžadují automatizaci.

Podle průzkumu provedeného Kanadskou stavební asociací a společností KPMG Kanada v roce 2021 [37] se pouze 25% společností domnívá, že jsou ve srovnání s konkurencí ve významném nebo odlišném postavení, pokud jde o zavádění technologií nebo digitálních řešení. Pouze 23% respondentů uvedlo, že jejich řešení jsou významně nebo ve velké míře založena na datech. Zároveň většina respondentů průzkumu charakterizovala své využívání řady dalších technologií jako čistě experimentální nebo přiznala, že je vůbec nepoužívá.

Tato neochota podílet se na technologických experimentech se projevuje zejména u velkých infrastrukturních projektů, kde chyby mohou stát miliony dolarů. I ty nejpokročilejší technologie - digitální dvojčata, prediktivní analytika - často narázejí na odpor nikoliv kvůli své efektivitě, ale kvůli nedostatečné spolehlivosti ověřené v reálných projektech.

Podle zprávy Světového ekonomického fóra (WEF) "Formování budoucnosti stavebnictví" [5] naráží zavádění nových technologií ve stavebnictví nejen na technické obtíže, ale také na psychologické bariéry ze strany zákazníků. [5] se zavádění nových technologií ve stavebnictví potýká nejen s technickými obtížemi, ale také s psychologickými bariérami na straně zákazníků. Mnozí zákazníci se obávají, že použití pokročilých řešení udělá z jejich projektů experimentální pracoviště a udělá z nich "pokusné králíky" a nepředvídatelné důsledky mohou vést k dalším nákladům a rizikům.



Obr. 2.1-11 Pro každý případ použití dat nabízí trh s řešeními aplikace pro optimalizaci a automatizaci procesů.

Stavebnictví je velmi rozmanité: různé projekty mají různé požadavky, regionální zvláštnosti, zákonné klasifikační předpisy (obr. 4.2-10), výpočtové normy (obr. 5.1-7) atd. Proto je prakticky nemožné vytvořit vlastní univerzální aplikaci nebo systém, který by dokonale vyhovoval všem těmto požadavkům a specifikům projektů.

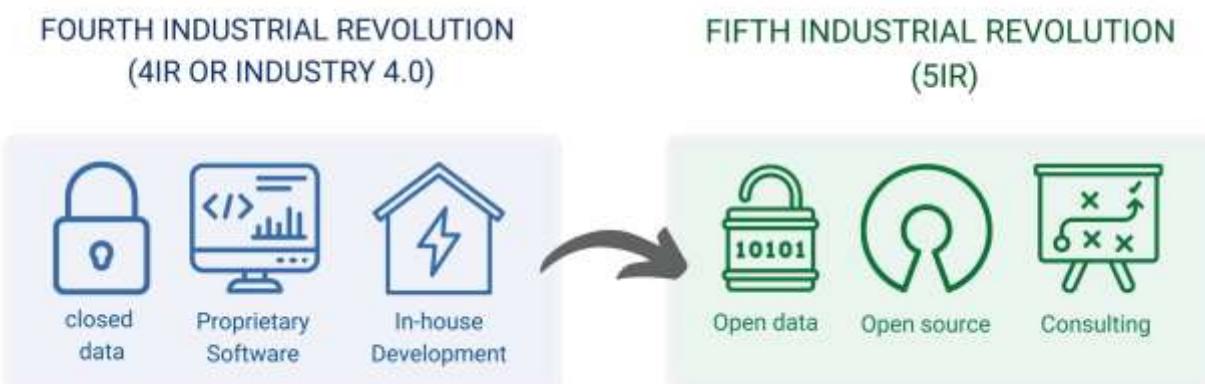
Ve snaze vypořádat se s rostoucí složitostí systémů a závislostí na dodavatelích softwaru se stále více ukazuje, že klíčem k efektivní správě dat je nejen otevřenosť a standardizace, ale také zjednodušení samotné architektury procesů. Rostoucí složitost a dynamika podnikových procesů vyžaduje nové přístupy, kdy se priorita přesouvá od hromadění dat k jejich strukturování a organizaci. Právě tento posun bude dalším krokem ve vývoji stavebnictví, který bude znamenat konec éry dominance dodavatelů softwaru a začátek éry smysluplné organizace informací.

Uvědomení si omezení univerzálních řešení a zranitelnosti vůči rostoucí složitosti vede k posunu priorit od uzavřených platforem a hromadění dat k transparentnosti, přizpůsobivosti a strukturovanému zpracování informací. Tento posun v myšlení odráží širší změny v globální ekonomice a technologiích, popisované optikou takzvaných "průmyslových revolucí". Abychom pochopili, kam stavebnictví směřuje a jakým směrem se bude ubírat, je třeba zvážit místo tohoto odvětví v kontextu čtvrté a páté průmyslové revoluce - od automatizace a digitalizace k personalizaci, otevřeným standardům a datovému modelu založenému na službách.

Čtvrtá průmyslová revoluce (Průmysl 4.0) a pátá průmyslová revoluce (Průmysl 5.0) ve stavebnictví.

Technologické a ekonomické etapy jsou teoretické koncepty používané k popisu a analýze vývoje společnosti a ekonomiky v různých fázích vývoje. Různí výzkumníci a odborníci je mohou interpretovat různě.

- **Čtvrtá průmyslová revoluce** (4IR nebo Průmysl 4.0) souvisí s informačními technologiemi, automatizací, digitalizací a globalizací. Jedním z jejích klíčových prvků je vytváření vlastních softwarových řešení, tj. specializovaných digitálních produktů určených pro konkrétní úkoly a podniky. Tato řešení se často stávají důležitou součástí IT infrastruktury, ale bez dalších úprav jsou špatně škálovatelná.
- **Pátá průmyslová revoluce** (5IR) je nyní v ranější fázi koncepce a vývoje než 4IR. Mezi její hlavní zásady patří větší personalizace výrobků a služeb. 5IR je hnutím směrem k přizpůsobivější, flexibilnější a personalizovanější hospodářské činnosti se zaměřením na personalizaci, poradenství a modely orientované na služby. Klíčovým aspektem pátého ekonomického módu je využívání dat pro rozhodování, které je prakticky nemožné bez využívání otevřených dat a otevřených nástrojů (obrázek 2.1-12).



Obr. 2.1-12 Čtvrtý vzor se zaměřuje na řešení, zatímco pátý vzor na personalizaci a data.

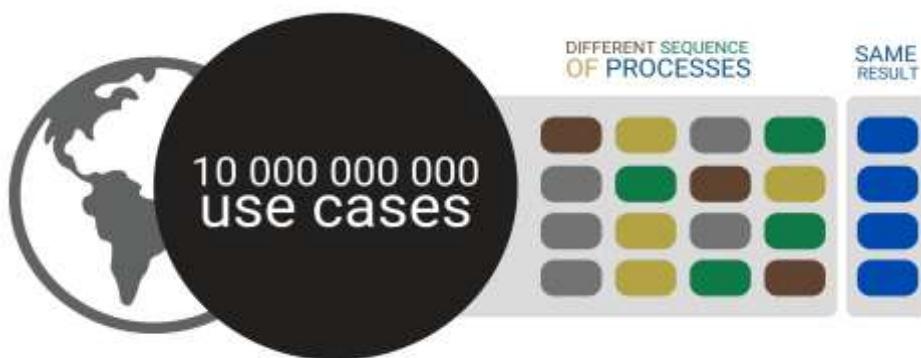
Vytvoření aplikace pro společnosti ve stavebnictví pro použití v deseti nebo sto organizacích nezaručuje její úspěšné rozšíření do dalších společností, regionů nebo zemí bez významných úprav a vylepšení. Pravděpodobnost úspěšného škálování takových řešení zůstává nízká, protože každá organizace má jedinečné procesy, požadavky a podmínky, které mohou vyžadovat individuální úpravy.

Je důležité si uvědomit, že úspěšná integrace technologických řešení již dnes znamená hluboce individuální přístup ke každému procesu, projektu a společnosti. To znamená, že i po vytvoření univerzálního rámce, nástroje nebo programu bude vyžadovat podrobné přizpůsobení a přizpůsobení tak, aby vyhovoval jedinečným požadavkům a podmínek každé konkrétní společnosti a projektu.

Podle zprávy PwC "Decoding the Fifth Industrial Revolution" [38]. [38] se letos přibližně 50% vrcholových manažerů v různých průmyslových odvětvích spoléhá na integraci pokročilých technologií a lidských odborných znalostí. Tento přístup jim umožňuje rychle se přizpůsobovat změnám v designu výrobků nebo požadavkům zákazníků a vytvářet individualizovanou výrobu.

Každý proces vyžaduje vývoj jedinečné funkce nebo aplikace, což vzhledem k velikosti světového stavebnictví a rozmanitosti projektů vede k existenci obrovského množství obchodních případů, které pokaždé představují jedinečnou logiku potrubí (obr. 2.1-13). Každý takový případ má svá specifika a vyžaduje individuální přístup. Podrobněji se rozmanitostí možných řešení téhož analytického problému v kontextu různých přístupů budeme zabývat v kapitole věnované strojovému učení a analýze datové sady Titanic (obr. 9.2-9).

Pipeline v kontextu digitálních procesů je posloupnost činností, procesů a nástrojů, které umožňují automatizovaný nebo strukturovaný tok dat a práce v různých fázích životního cyklu projektu.



Obr. 2.1-13 Individuálnost a variabilita obchodních případů znemožňuje pokusy o vytvoření škálovatelných uzavřených platform a nástrojů.

Vlivem digitální transformace se náš život již v mnoha ohledech změnil a dnes můžeme hovořit o nástupu nové etapy v ekonomickém vývoji stavebnictví. V této "nové ekonomice" bude soutěž organizována podle jiných pravidel: kdo dokáže efektivně proměnit veřejné znalosti a otevřená data v žádané produkty a služby, získá v podmírkách páté průmyslové revoluce klíčovou výhodu.

Jak uvádí ekonomka Kate Maskus v knize "Private Rights and Public Problems: The Global Intellectual Property Economy in the 21st Century" [39]. 2012 [39], "žijeme v globální znalostní ekonomice a budoucnost patří těm, kteří vědí, jak přeměnit vědecké objevy na zboží".

Přechod na pátý ekonomický režim znamená posun od uzavřených IT řešení k otevřeným standardům a platformám. Společnosti se začnou odklánět od tradičních softwarových produktů ve prospěch modelů orientovaných na služby, kde se hlavním aktivem stanou data, nikoliv proprietární technologie.

Studie Harvard Business School 2024 [40] ukazuje obrovskou ekonomickou hodnotu otevřeného softwaru (Open Source Software, OSS). Podle této studie je OSS přítomen v 96% všech softwarových kódů a některé komerční softwary se z 99,9% skládají z komponent OSS. Bez OSS by společnosti utratily za software 3,5krát více.

Budování podnikových ekosystémů bude v souladu s globálními trendy postupně přecházet do pátého ekonomického paradigmatu, v němž budou mít větší prioritu datově orientované analytické a poradenské služby než izolovaná, uzavřená řešení s pevně definovanými scénáři použití.

Éra digitalizace změní poměr sil v odvětví: namísto spoléhání se na řešení od dodavatelů budou společnosti zakládat svou konkurenceschopnost na schopnosti efektivně využívat data. V důsledku toho se stavebnictví posune od starých rigidních systémů k flexibilním, adaptivním ekosystémům, kde základem řízení projektů budou otevřené standardy a interoperabilní nástroje. Konec éry dominance dodavatelů aplikací vytvoří nové prostředí, v němž hodnotu nebude definovat vlastnictví uzavřeného zdrojového kódu a specializovaných konektorů, ale schopnost proměnit data ve strategickou výhodu.



KAPITOLA 2.2.

PŘEMĚNA CHAOSU V ŘÁD A SNÍŽENÍ SLOŽITOSTI

Zbytečný kód a uzavřené systémy jako překážka zvýšení produktivity

V uplynulých desetiletích byly technologické změny v oblasti IT řízeny především dodavateli softwaru. Ti určují směr vývoje a určují, které technologie mají společnosti přjmout a které mají být ponechány stranou. V éře přechodu od oddělených řešení k centralizovaným databázím a integrovaným systémům prosazovali dodavatelé licencované produkty, které poskytovaly kontrolu nad přístupem a škálovatelností. Později, s příchodem cloudových technologií a modelů softwaru jako služby (SaaS), se tato kontrola vyvinula v model předplatného, což upevnilo uživatele jako loajální zákazníky digitálních služeb.

Tento přístup vedl k paradoxu: navzdory nebývalému objemu vytvořeného programového kódu je jen malá část z něj skutečně využívána. Možná je kódu stokrát či tisíckrát více, než je nutné, protože stejně podnikové procesy jsou popsány a duplikovány v desítkách či stovkách programů různými způsoby, a to i v rámci jedné firmy. Zároveň již byly zaplaceny náklady na vývoj, které jsou nevratné. Přesto průmysl pokračuje v reprodukci tohoto cyklu a vytváří nové produkty s minimální přidanou hodnotou pro koncového uživatele, častěji pod tlakem očekávání trhu než skutečných potřeb.

Podle příručky Defence Acquisition University (DAU) Software Development Cost Estimating Guide [41] se náklady na vývoj softwaru mohou výrazně lišit v závislosti na několika faktorech, včetně složitosti systému a zvolené technologie. Historicky se náklady na vývoj v roce 2008 pohybovaly kolem 100 USD na rádek zdrojového kódu (SLOC), zatímco náklady na údržbu se mohou vyšplhat až na 4 000 USD na SLOC.

Jen jedna z komponent aplikací CAD - geometrické jádro - může mít desítky milionů řádků kódu (obr. 6.1-5). Podobná situace je i v systémech ERP (obr. 5.4-4), k jejichž diskusi o složitosti se vrátíme v páté části knihy. Při bližším pohledu však zjistíme, že většina tohoto kódu nepřináší přidanou hodnotu, ale funguje pouze jako "pošták" - mechanicky přesouvá data mezi databází, rozhraním API, uživatelským rozhraním a dalšími tabulkami v systému. Navzdory rozšířenému mýtu o zásadním významu takzvané obchodní logiky je drsná realita mnohem prozaicitější: moderní kódové báze jsou plné zastaralých šablonových bloků (legacy code), jejichž jediným účelem je zajistit přenos dat mezi tabulkami a komponentami, aniž by ovlivnily rozhodování nebo efektivitu podnikání.

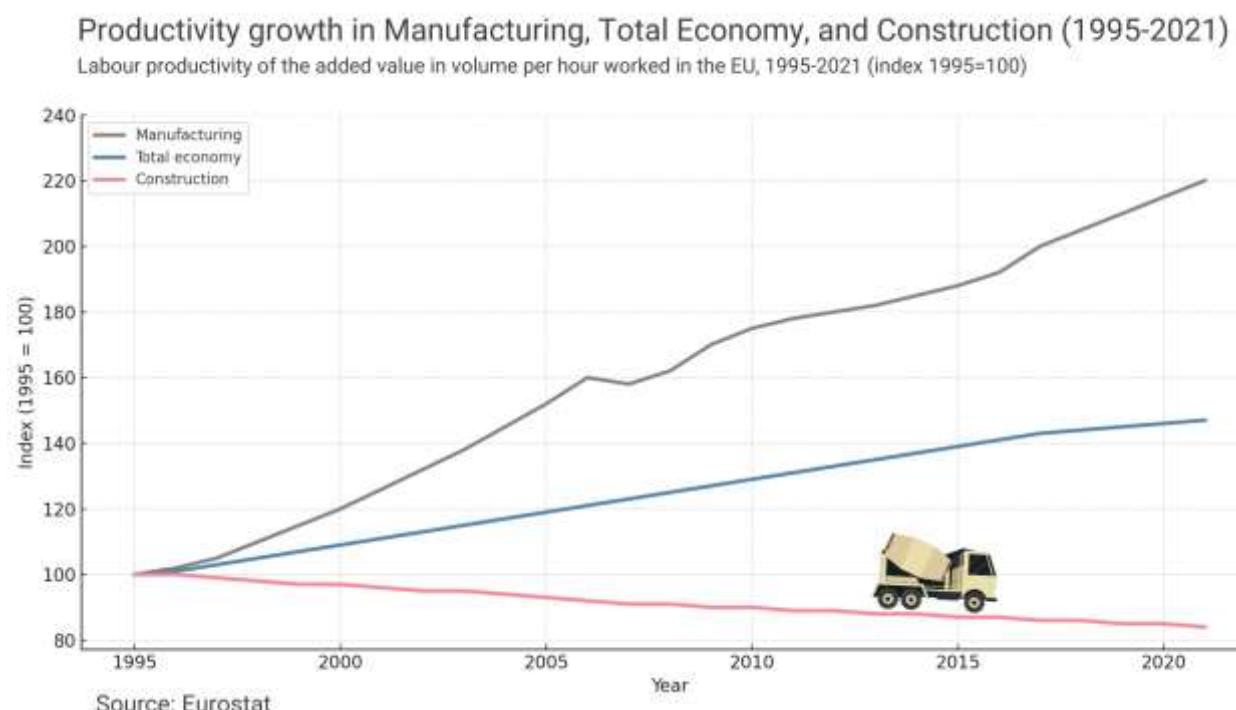
V důsledku toho se uzavřená řešení, která zpracovávají data z různých zdrojů, nevyhnutelně mění v nepřehledné "špagetové ekosystémy". Tyto složité, vzájemně provázané systémy může zvládnout pouze armáda manažerů pracujících v poloprovozním režimu. Taková organizace správy dat je nejen neefektivní z hlediska zdrojů, ale také vytváří kritické slabiny v podnikových procesech a činí podnik závislým na úzkém okruhu specialistů, kteří rozumí fungování tohoto technologického bludiště.

Neustálý nárůst objemu kódu, počtu aplikací a rostoucí složitost konceptů nabízených dodavateli vedly k přirozenému výsledku - nárůstu složitosti ekosystému IT ve stavebnictví. To způsobilo, že praktická implementace digitalizace prostřednictvím zvyšování počtu aplikací v oboru je neefektivní. Softwarové produkty vytvořené bez náležitého zohlednění potřeb uživatelů často vyžadují značné

prostředky na implementaci a podporu, ale nepřináší očekávanou návratnost.

Podle studie společnosti McKinsey "Zvyšování produktivity ve stavebnictví" [42] činil v posledních dvou desetiletích celosvětový růst produktivity práce ve stavebnictví v průměru pouze 1% ročně ve srovnání s růstem 2,8% v celé světové ekonomice a 3,6% ve zpracovatelském průmyslu. Ve Spojených státech klesla produktivita práce ve stavebnictví na pracovníka od 60. let 20. století na polovinu [43].

Rostoucí složitost systémů, izolace a uzavřenosť dat zhorsily komunikaci mezi odborníky, čímž se stavebnictví stalo jedním z nejméně efektivních (obrázek 2.2-1). na 22 bilionů dolarů do roku 2040, což bude vyžadovat výrazné zvýšení efektivity.



Obr. 2.2-1 Uzavřenosť a složitosť dat a v důsledku toho špatná komunikace mezi odborníky vedly k tomu, že se stavebnictví stalo jedním z nejméně efektivních odvětví ekonomiky (podle [44], [45]).

Jak zdůrazňuje studie společnosti McKinsey (2024) "Zajištění produktivity ve stavebnictví již není volitelné", s rostoucím nedostatkem zdrojů a snahou odvětví zdvojnásobit tempo růstu si stavebnictví již nemůže dovolit zůstat na současné úrovni produktivity [44].

Předpokládá se, že celosvětové náklady na stavebnictví vzrostou z 13 bilionů dolarů v roce 2023 na mnohem vyšší úroveň do konce desetiletí, takže otázka efektivity je nejen aktuální, ale kritická.

Jedním z klíčových způsobů zvýšení efektivity bude nevyhnutelné sjednocení a zjednodušení aplikačních struktur a architektury datových ekosystémů. Tento přístup k racionalizaci odstraní nadbytečné vrstvy abstrakce a zbytečnou složitost, které se v podnikových systémech za léta nahromadily.

Od silosů k jednotnému datovému skladu

Čím více dat organizace hromadí, tím obtížnější je z nich získat skutečnou hodnotu. Vzhledem k roztríštěnosti ukládání informací v izolovaných silech jsou obchodní procesy moderních společností podobné stavitelům, kteří se snaží postavit mrakodrap z materiálů uložených v tisících různých skladů. Nadbytek informací nejenže ztěžuje přístup k právně relevantním informacím, ale také zpomaluje rozhodování: každý krok je třeba opakován kontrolou a potvrzovat.

Každá úloha nebo proces je pevně spojen se samostatnou tabulkou nebo databází a výměna dat mezi systémy vyžaduje složité integrace. Chyb a nekonzistence v jednom systému mohou způsobit řetězová selhání v ostatních. Nesprávné hodnoty, opožděné aktualizace a duplicitní informace nutí zaměstnance trávit značný čas ručním porovnáváním a sladováním dat. V důsledku toho organizace tráví více času řešením důsledků roztríštěnosti než vývojem a optimalizací procesů.

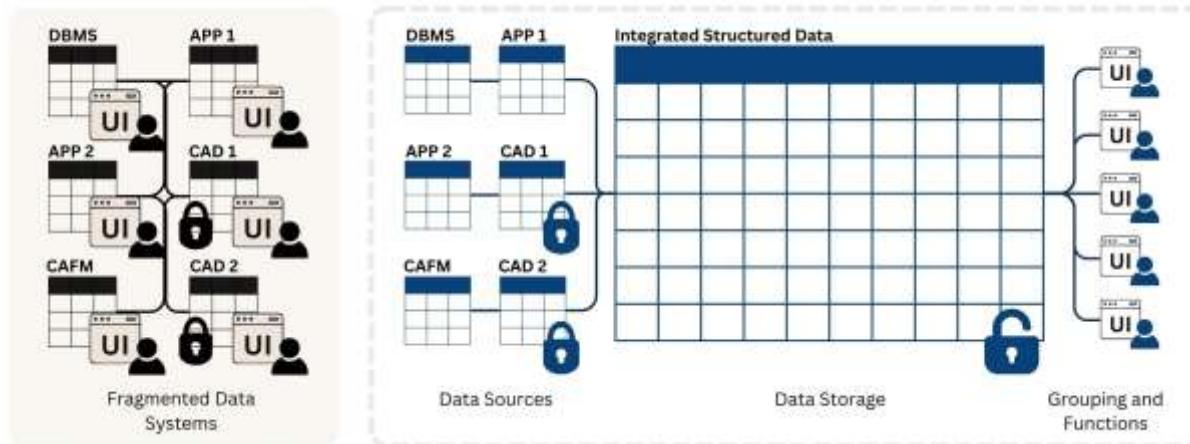
Tento problém je univerzální: některé společnosti se stále potýkají s chaosem, zatímco jiné nacházejí řešení v integraci - přesunu informačních toků do centralizovaného systému ukládání. Představte si jej jako jednu velkou tabulku, do které můžete ukládat veškeré entity související s úkoly, projekty a objekty. Místo desítek různorodých tabulek a formátů vznikne jediné ucelené úložiště (obrázek 2.2-2), které umožňuje:

- minimalizovat ztrátu dat;
- odstranit potřebu neustálé harmonizace informací;
- zlepšit dostupnost a kvalitu dat;
- zjednodušit analytické zpracování a strojové učení

Převedení dat do společného standardu znamená, že bez ohledu na zdroj jsou informace převedeny do jednotného a strojově čitelného formátu. Takové uspořádání dat umožňuje kontrolovat jejich integritu, analyzovat je v reálném čase a pohotově je využívat pro manažerská rozhodnutí.

Koncepci integrovaných úložných systémů a jejich využití v analytice a strojovém učení se budeme podrobněji věnovat v kapitole "Úložiště velkých dat a strojové učení". Tématům modelování a strukturování dat se budeme podrobně věnovat v kapitolách "Transformace dat do strukturované podoby"

a "Jak standardy mění hru: od náhodných souborů k propracovanému datovému modelu".



Obr. 2.2-2 Datová integrace odstraňuje "sila", zlepšuje dostupnost informací a optimalizuje podnikové procesy.

Po strukturování a sloučení dat je dalším logickým krokem jejich validace. Díky jedinému integrovanému úložišti se tento proces výrazně zjednoduší: už žádné vícenásobné nekonzistentní schéma, duplicitní struktury a složité vztahy mezi tabulkami. Všechny informace jsou sladěny s jediným datovým modelem, což eliminuje vnitřní nekonzistence a urychluje proces validace. Validace a zajištění kvality dat jsou základními aspekty všech podnikových procesů a podrobněji se jim budeme věnovat v příslušných kapitolách knihy.

V závěrečné fázi jsou data seskupena, filtrována a analyzována. Jsou na ně aplikovány různé funkce: agregace (sčítání, násobení), výpočty mezi tabulkami, sloupci nebo řádky (obr. 2.2-4). Práce s daty se stává posloupností kroků: sběr, strukturování, validace, transformace, analytické zpracování a předání do finálních aplikací, kde se informace využívají k řešení praktických problémů. Více o sestavování takových scénářů, automatizaci kroků a sestavování toků zpracování si povíme v kapitolách o ETL - procesech a přístupu k datovému potrubí.

Digitální transformace tedy není jen o zjednodušení práce s informacemi. Jde o odstranění nadmerné složitosti při správě dat, o přechod od chaosu k předvídatelnosti, od mnoha systémů ke zvládnutelnému procesu. Čím nižší je složitost architektury, tím méně kódu je potřeba k její podpoře. A v budoucnu může kód jako takový zmizet a ustoupit inteligentním agentům, kteří budou nezávisle analyzovat, systematizovat a transformovat data.

Integrované úložné systémy umožňují přechod na agenty s umělou inteligencí

Čím méně složitá jsou data a systémy, tím méně kódu je třeba napsat a udržovat. A nejjednodušší způsob, jak ušetřit vývoj, je zbavit se kódu úplně a nahradit ho daty. Když se vývoj kódu aplikací přesune od kódu k datovým modelům, nevyhnutelně dojde k posunu směrem k přístupu zaměřenému na data (data-driven), protože za těmito koncepty stojí zcela jiný způsob myšlení.

Když se člověk rozhodne pracovat s daty v centru, začne jejich roli vnímat jinak. Data už nejsou jen "surovinou" pro aplikace - nyní jsou základem, kolem kterého se buduje architektura, logika a interakce.

Tradiční přístup ke správě dat obvykle začíná na úrovni aplikace a v konstrukci připomíná těžkopádný byrokratický systém: víceúrovňové schvalování, ruční kontroly, nekonečné verze dokumentů prostřednictvím příslušných softwarových produktů. S rozvojem digitálních technologií bude stále více firem nuteno přejít na princip minimalismu - ukládat a používat jen to, co je skutečně nutné a bude se používat.

Logiku minimalizace převzali prodejci. V zájmu zjednodušení ukládání a zpracování dat se práce uživatelů přesouvá z offline aplikací a nástrojů do cloudových služeb a takzvaných řešení SaaS.

Koncept SaaS (Software as a Service, neboli "software jako služba") je jedním z klíčových trendů v moderních IT infrastrukturách, který umožňuje uživatelům přistupovat k aplikacím prostřednictvím internetu, aniž by museli instalovat a udržovat software na vlastních počítačích.

Na jedné straně SaaS usnadnilo škálování, řízení verzí a snížilo náklady na podporu a údržbu, ale na druhé straně vedle závislosti na logice konkrétní aplikace také způsobilo, že uživatel je zcela závislý na cloudové infrastruktuře poskytovatele. V případě výpadku služby může být přístup k datům a podnikovým procesům dočasně nebo dokonce trvale zablokován. Kromě toho jsou všechna uživatelská data při práci s aplikacemi SaaS uložena na serverech poskytovatele, což vytváří bezpečnostní rizika a rizika související s dodržováním předpisů. Změny tarifů nebo podmínek používání mohou také vést ke zvýšení nákladů nebo k nutnosti urychlené migrace.

Vývoj umělé inteligence, LLM -agentů a přístupu zaměřeného na data zpochybnil budoucnost aplikací v jejich tradiční podobě a provedení SaaS. Zatímco dříve bylo nutné, aby aplikace a služby spravovaly obchodní logiku a zpracovávaly data, s nástupem agentů AI se tyto funkce mohou přesunout na inteligenční systémy, které pracují přímo s daty.

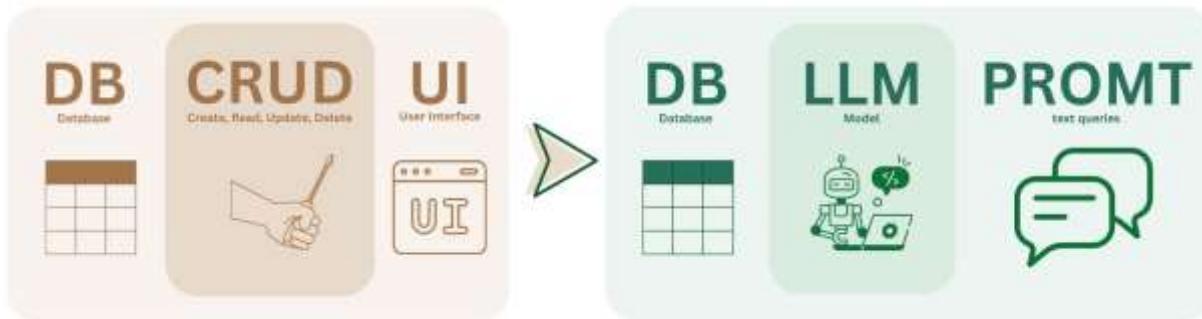
Proto se v IT odděleních a na úrovni managementu stále častěji hovoří o hybridních architekturách, v nichž se cloudové služby doplňují s AI -agenty a lokálními řešeními, což snižuje závislost na platformách SaaS -platformách.

Náš přístup uznává, že tradiční podnikové aplikace nebo aplikace SaaS se mohou v době agentů výrazně změnit. Tyto aplikace jsou v podstatě CRUD [create, read, update and delete] databáze s obchodní logikou. V budoucnu však tuto logiku převezmou agenti s umělou inteligencí [46].

- Satya Nadella, generální ředitel společnosti Microsoft, 2024.

Přístup zaměřený na data a využití agentů AI/LLM může omezit nadbytečné procesy, a tím snížit pracovní zátěž zaměstnanců. Pokud jsou data správně uspořádána, je snazší je analyzovat, vizualizovat a použít při rozhodování. Místo nekonečných reportů a kontrol získají specialisté přístup k aktuálním informacím na několik kliknutí nebo s pomocí agentů LLM automaticky v podobě hotových dokumentů a dashboardů.

Při manipulaci s daty nám budou pomáhat nástroje umělé inteligence (AI) a chaty LLM. V posledních letech je patrný trend odklonu od tradičních operací CRUD (create, read, update, delete) směrem k využívání velkých jazykových modelů (LLM) pro správu dat. LLM jsou schopny interpretovat přirozený jazyk a automaticky generovat vhodné databázové dotazy, což zjednodušuje interakci se systémy pro správu dat (obr. 2.2-3).



Obr. 2.2-3 AI nahradí a integruje úložná a databázová řešení a postupně vytlačí tradiční aplikace a CRUD -operace.

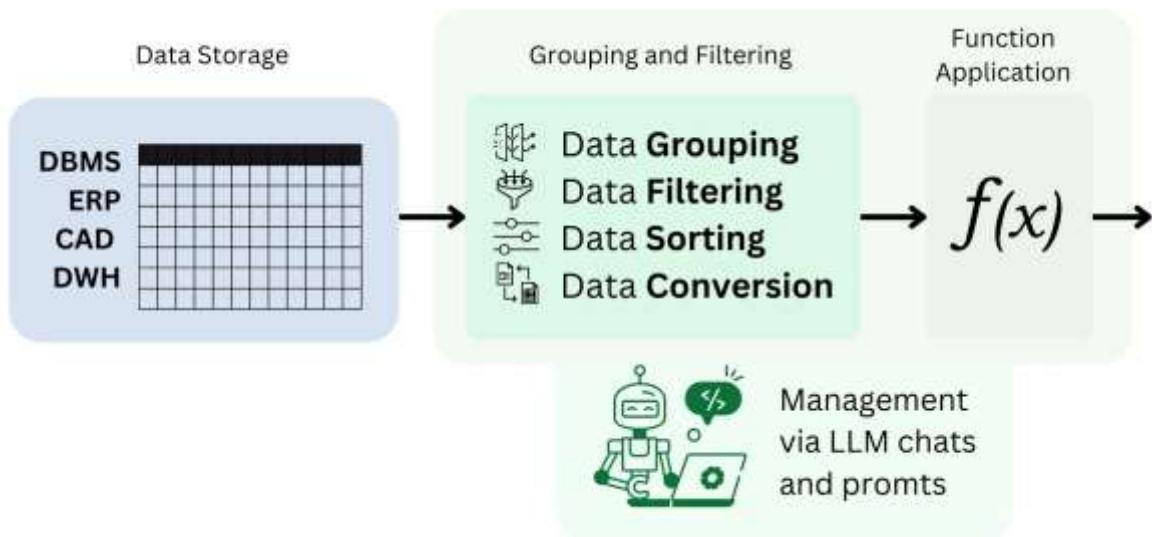
V příštích 3-6 měsících bude 90% kódu psát umělá inteligence a za 12 měsíců by téměř veškerý kód mohla generovat umělá inteligence [47].

- Dario Amodei, generální ředitel společnosti LLM Anthropic, březen 2025.

Navzdory rychlému rozvoji nástrojů pro vývoj umělé inteligence (např. GitHub Copilot) budou v roce 2025 v tomto procesu stále hrát klíčovou roli vývojáři. Agenti AI se stávají stále užitečnějšími pomocníky: automaticky interpretují dotazy uživatele, generují dotazy SQL a Pandas (více o tom v následujících kapitolách) nebo píší kód pro analýzu dat. Umělá inteligence tak postupně nahrazuje tradiční uživatelská rozhraní aplikací.

Rozšiřování modelů umělé inteligence, jako jsou jazykové modely, bude hnacím motorem vývoje hybridních architektur. Namísto úplného opuštění cloudových řešení a produktů SaaS můžeme být svědky integrace cloudových služeb s lokálními systémy správy dat. Například federativní učení umožňuje výkonné modely umělé inteligence bez nutnosti přesouvat citlivá data do cloutu. Firmy si

tak mohou zachovat kontrolu nad svými daty a zároveň získat přístup k pokročilým technologiím.



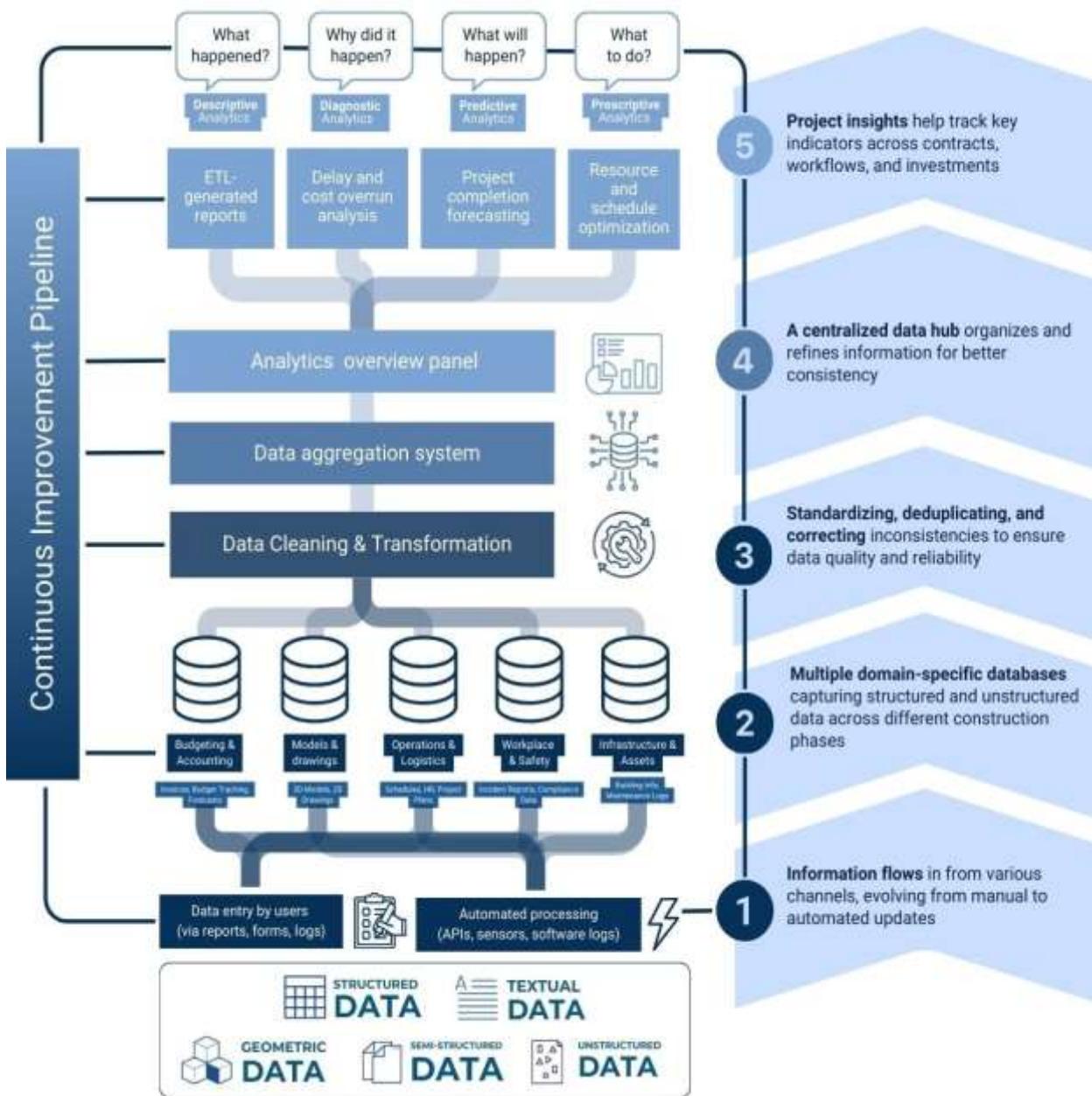
Obrázek 2.2-4 Základní operace seskupování, filtrování a třídění s následnou aplikací funkcí budou zpracovávat chaty LLM.

Budoucnost stavebnictví bude založena na kombinaci lokálních řešení, cloudového výkonu a inteligentních modelů, které společně vytvářejí efektivní a bezpečné systémy správy dat. LLM umožní uživatelům bez hlubokých technických znalostí komunikovat s databázemi a datovými sklady formulováním dotazů v přirozeném jazyce. O agentech LLM a AI a jejich fungování si povíme více v kapitole "Agenti LLM a formáty strukturovaných dat".

Správně uspořádaná data a jednoduché, snadno použitelné analytické nástroje s podporou LLM nejen usnadní práci s informacemi, ale také pomohou minimalizovat chyby, zvýšit efektivitu a automatizovat procesy.

Od sběru dat k rozhodování: cesta k automatizaci

V dalších částech knihy se budeme podrobně zabývat tím, jak spolu specialisté komunikují a jak se data stávají základem pro rozhodování, automatizaci a provozní efektivitu. Na obrázku 2.2-5 je uveden příklad diagramu znázorňujícího posloupnost kroků zpracování dat v datově orientovaném přístupu. Tento diagram znázorňuje kontinuální zlepšovací potrubí, jehož části budou podrobně rozebrány později v knize.



Obr. 2.2-5 Příklad kontinuálního zlepšování dat: tok zpracování a analýzy dat ve stavebních projektech.

Systém popisující obchodní procesy středně velké společnosti je postaven na víceúrovňovém principu. Zahrnuje: sběr dat, čištění, agregaci, analytické zpracování a rozhodování na základě výsledků. Všem těmto fázím se budeme věnovat dále v knize - jak v teoretickém kontextu, tak na praktických příkladech:

- Na první úrovni probíhá **zadávání údajů** (obr. 3.1-1). Informace jsou přijímány jak manuálně (prostřednictvím hlášení, formulářů, protokolů), tak v automatizované podobě (z rozhraní API, senzorů, softwarových systémů). Data mohou mít různou strukturu: geometrickou, textovou,

nestrukturovanou. V této fázi je potřeba standardizovat, strukturovat a sjednotit informační toky.

- Další úrovní je **zpracování a transformace dat**. Zahrnuje procesy čištění, odstraňování duplicit, opravy chyb a přípravy informací pro další analýzu (obrázek 4.2-5). Tato fáze je kriticky důležitá, protože kvalita analytických dat přímo závisí na čistotě a přesnosti dat.
- Data pak proudí do **specializovaných tabulek, datových rámců nebo databází** rozdelených podle funkčních oblastí: rozpočet a účetnictví, modely a výkresy, logistika, bezpečnost a infrastruktura. Toto rozdelení umožňuje snadný přístup k informacím a jejich vzájemné analýze.
- Data jsou poté **agregována a zobrazena na analytickém panelu** (showcase). Zde se uplatňují metody deskriptivní, diagnostické, prediktivní a preskriptivní analýzy. To pomáhá odpovědět na klíčové otázky (obrázek 1.1-4): co se stalo, proč se to stalo, co se stane v budoucnu a jaká opatření je třeba přijmout. Systém může například identifikovat zpoždění, předvídat dokončení projektů nebo optimalizovat zdroje.
- A konečně poslední úroveň vytváří **analytické závěry a klíčové ukazatele**, které pomáhají sledovat plnění smluv, řídit investice a zlepšovat podnikové procesy (obr. 7.4-2). Tyto informace se stávají základem pro rozhodování a strategii rozvoje podniku.

Stejně tak se data od sběru dostávají k využití ve strategickém řízení. V následujících částech knihy se budeme podrobně zabývat jednotlivými fázemi a zaměříme se na typy dat, techniky jejich zpracování, analytické nástroje a reálné případy využití těchto přístupů ve stavebnictví.

Další kroky: přeměna chaosu na zvládnutelný systém

V této části jsme zkoumali problémy informačních sil a zabývali se dopadem přílišné složitosti systému na výkonnost podniku a analyzovali přechod od čtvrté průmyslové revoluce k páté, kde jsou ústředním tématem data, nikoli aplikace. Viděli jsme, jak oddělené informační systémy vytvářejí překážky pro sdílení znalostí a jak pokračující složitost IT prostředí snižuje produktivitu a brzdí inovace ve stavebnictví.

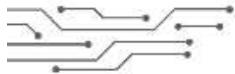
Pro shrnutí této části je vhodné zdůraznit hlavní praktické kroky, které vám pomohou aplikovat diskutované přístupy ve vašich každodenních úkolech:

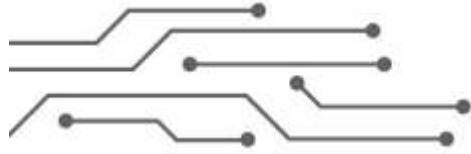
- Vizualizace informačního prostředí
 - Vytvořte si vizuální mapu zdrojů dat (Miro, Figma, Canva), se kterými pravidelně pracujete.
 - Přidejte do této mapy systémy a aplikace, které používáte při své práci.
 - Identifikace potenciálně duplicitních funkcí a nadbytečných řešení.
 - Identifikace kritických míst, kde může dojít ke ztrátě nebo poškození dat během přenosu mezi systémy.
- Zavedení personalizovaných postupů správy dat
 - Přesun pozornosti od aplikací k datům jako klíčovému aktivu v procesech.
 - Zdokumentování zdrojů dat a metodiky zpracování pro zajištění transparentnosti

- Vyvinout mechanismy pro hodnocení a zlepšování kvality dat
 - Snažte se zajistit, aby se údaje zadávaly jednou a používaly opakovaně - to je základem efektivní organizace procesů.
- Podporujte ve svém týmu přístup zaměřený na data (data-driven).
- navrhnut používání standardizovaných a jednotných formátů pro vzájemnou výměnu dat.
 - Pravidelně na schůzkách týmu vznášet otázky týkající se kvality a dostupnosti dat.
 - Seznamte se s alternativami nástrojů Open Source, které používáte k řešení svých problémů.

Začněte v malém - vyberte si jeden konkrétní proces nebo soubor dat, který je pro vaši práci klíčový, a použijte na něj přístup zaměřený na data, přičemž se zaměříte z nástrojů na data. Dosažením úspěchu v jediném pilotním projektu získáte nejen praktické zkušenosti, ale také jasnou ukázkou přínosů nové metodiky pro váš tým. Při plnění většiny těchto kroků, pokud budete mít otázky, můžete požádat o vysvětlení a pomoc některý z aktuálních LLM.

V následujících částech knihy se budeme podrobněji zabývat technikami strukturování a harmonizace dat a praktickými přístupy k integraci heterogenních informací. Zvláštní pozornost bude věnována přechodu od nesourodých datových sil k jednotným datovým ekosystémům, které hrají klíčovou roli v digitální transformaci stavebnictví.





III ČÁST

DATOVÝ RÁMEC V PODNIKOVÝCH PROCESECH VE STAVEBNICTVÍ

Ve třetí části je vytvořen komplexní přehled o typologii dat ve stavebnictví a metodách jejich efektivní organizace. Jsou analyzovány charakteristiky a specifika práce se strukturovanými, nestrukturovanými, polostrukturovanými, textovými a geometrickými daty v kontextu stavebních projektů. Jsou rozebrány moderní formáty ukládání a protokoly pro výměnu informací mezi různými systémy používanými v oboru. Jsou popsány praktické nástroje a techniky pro převod dat různých formátů do jednotného strukturovaného prostředí, včetně způsobu integrace dat CAD (BIM). Jsou navrženy přístupy k zajištění kvality dat prostřednictvím standardizace a validace, které jsou rozhodující pro přesnost stavebních výpočtů. Podrobně jsou rozebrány praktické aspekty využití moderních technologií (Python Pandas, LLM - modely) s příklady kódu pro řešení typických problémů ve stavebnictví. Je zdůvodněn význam vytvoření kompetenčního centra (CoE) jako organizační struktury pro koordinaci a standardizaci přístupů k řízení informací.

KAPITOLA 3.1.

DATOVÉ TYPY VE STAVEBNICTVÍ

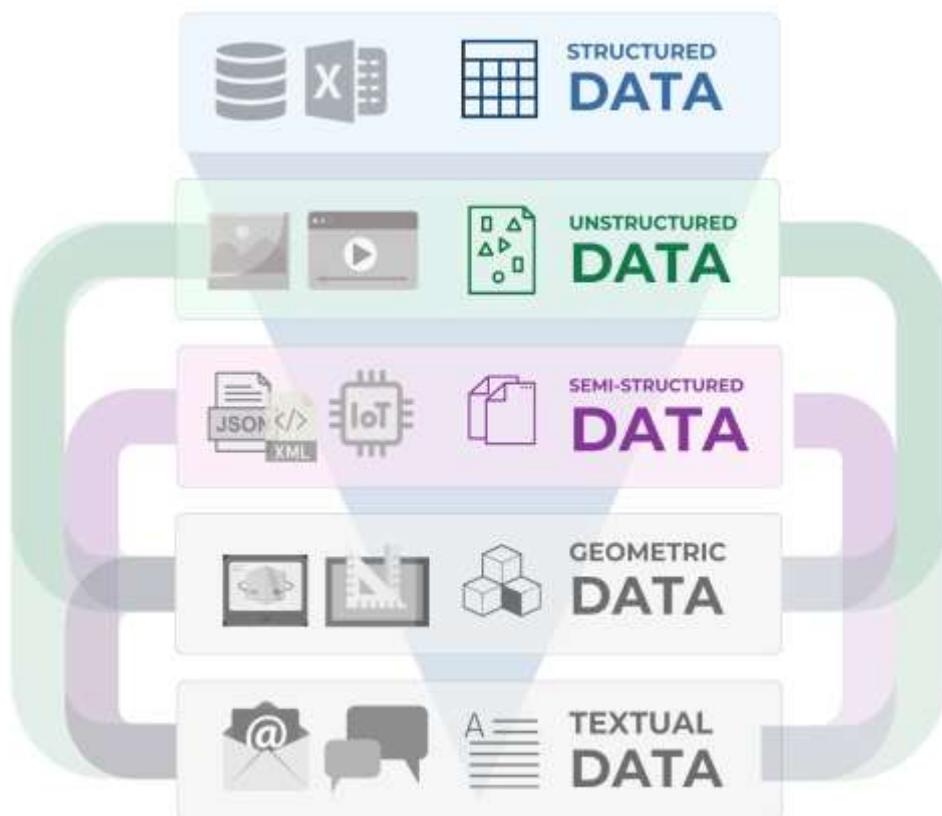
Nejdůležitější typy dat ve stavebnictví

V moderním stavebnictví se systémy, aplikace a datové skladby společností aktivně plní informacemi a daty různých typů a formátů (obr. 3.1-1). Podívejme se blíže na hlavní typy dat, které tvoří informační krajину moderní společnosti působící ve stavebnictví:

- **Strukturovaná** data: tato data mají jasnou organizační strukturu, např. tabulky Excelu a relační databáze.
- **Nestrukturovaná** data: jedná se o informace, které nejsou uspořádány podle přísných pravidel. Příkladem takových dat je text, video, fotografie a zvukové záznamy.
- **Volně strukturovaná** data: tato data zaujímají mezipolohu mezi strukturovanými a nestrukturovanými daty. Obsahují prvky struktury, ale tato struktura není vždy jasná nebo je často popsána prostřednictvím různých schémat. Příkladem polostrukturovaných dat ve stavebnictví jsou: technické specifikace, projektová dokumentace nebo zprávy o postupu prací.
- **Textová data**: zahrnují vše, co pochází z ústní a písemné komunikace, jako jsou e-maily, přepisy schůzek a jednání.
- **Geometrická** data: tato data pocházejí z programů CAD, ve kterých specialisté vytvářejí geometrická data prvků projektu pro vizualizaci, potvrzení objemových hodnot nebo kontrolu kolizí.

Je důležité poznamenat, že geometrické a textové (alfanumerické) údaje nejsou samostatnou kategorií, ale mohou se vyskytovat ve všech třech typech údajů. Například geometrická data, mohou být součástí jak strukturovaných dat (parametrické formáty CAD), tak nestrukturovaných dat (naskenované výkresy). Textová data mohou být podobně jak uspořádána v databázích (strukturovaná data), tak mohou existovat jako dokumenty bez jasné struktury.

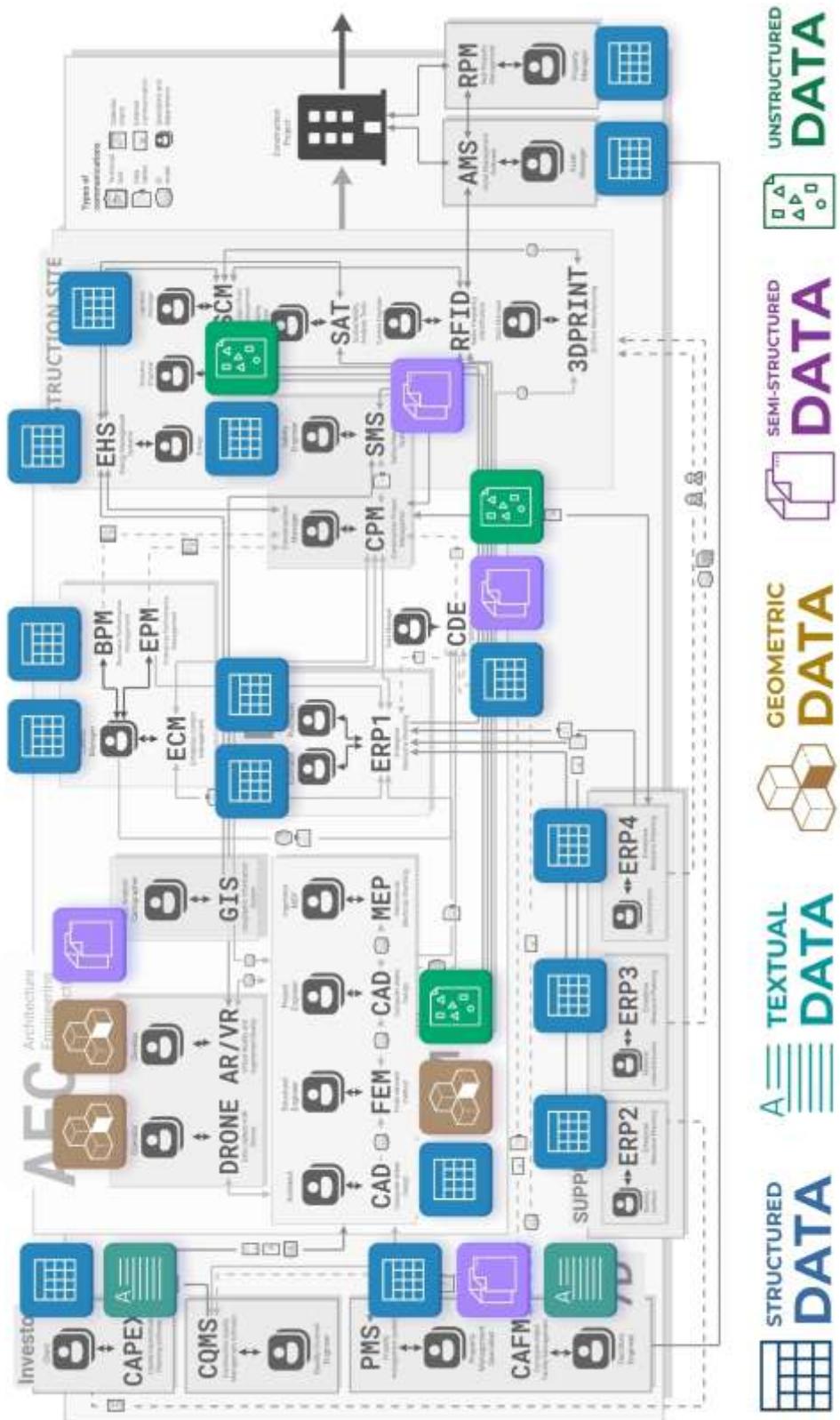
Každý typ dat ve stavební společnosti je jedinečným prvkem v mozaice informačních aktiv společnosti. Od nestrukturovaných dat, jako jsou obrázky ze stavenišť a zvukové záznamy z jednání, až po strukturované záznamy včetně tabulek a databází, každý prvek hraje důležitou roli při utváření informačního prostředí společnosti.



Obr. 3.1-1 Inženýři a správci dat se musí naučit pracovat se všemi typy dat používaných ve stavebnictví.

Zde je uveden ukázkový seznam jen některých systémů a souvisejících typů dat (obrázek 3.1-2) používaných ve stavebnictví:

- **ERP** (Enterprise Resource Planning) - zpracovává obecně strukturovaná data, která pomáhají spravovat podnikové zdroje a integrovat různé podnikové procesy.
- **CAD** (Computer-Aided Design) v kombinaci s **BIM** (Building Information Modeling) - využívá geometrická a polostrukturovaná data k navrhování a modelování stavebních projektů, čímž zajišťuje přesnost a konzistence informací ve fázi návrhu.
- **GIS** (Geografické informační systémy) - pracuje s geometrickými a strukturovanými daty a vytváří a analyzuje kartografická data a prostorové vztahy.
- **RFID** (Radio-Frequency Identification) - využívá polostrukturovaná data k efektivnímu sledování materiálů a zařízení na staveništi pomocí radiofrekvenční identifikace.
- **ECM** (Engineering Content Management) je systém pro správu technických dat a dokumentace, včetně polostrukturovaných a nestrukturovaných dat, jako jsou technické výkresy a konstrukční dokumenty.

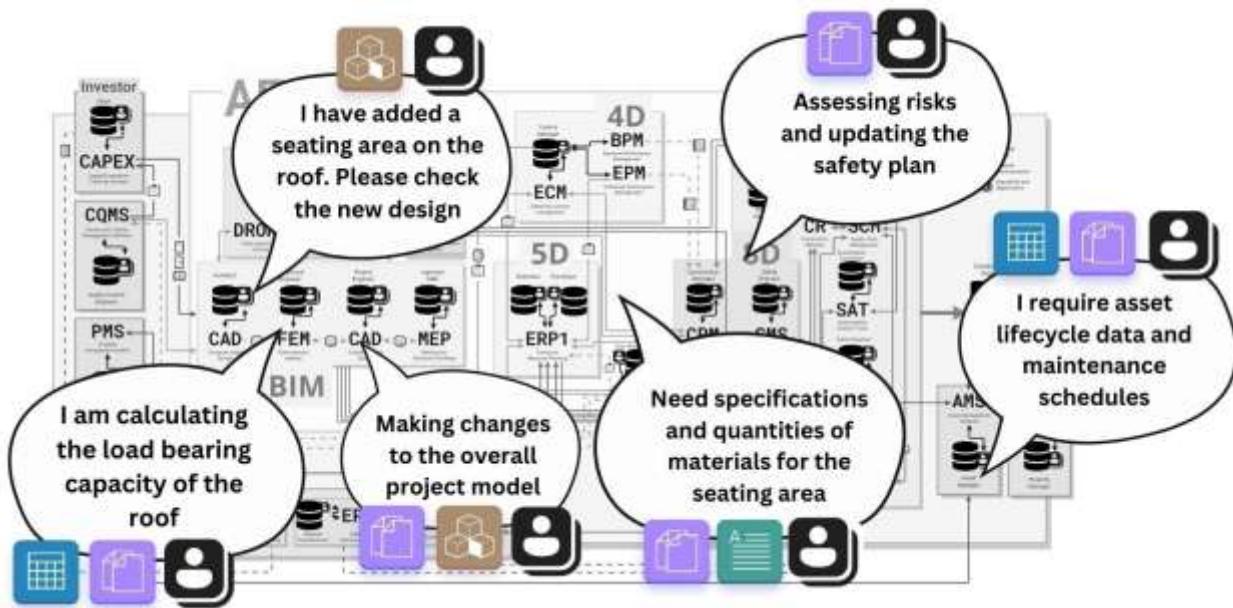


Obr. 3.1-2 Různé formáty a data vyplňují různé systémy, což vyžaduje překlad do podoby vhodné pro komplexní integraci.

Tyto a mnohé další systémy společnosti spravují širokou škálu dat, od strukturovaných tabulkových dat až po složité geometrické modely, a zajišťují integrovanou spolupráci v procesech navrhování, plánování a řízení výstavby.

V příkladu zjednodušeného dialogu (obr. 3.1-3) si specialisté stavebního projektu vyměňují různé typy údajů:

- ⦿ **Architekt:** "S ohledem na přání klienta jsem na střechu přidal posezení. Podívejte se prosím na nový návrh" (geometrické údaje - model).
- ⦿ **Stavební inženýr:** "Projekt byl přijat. Provádím výpočet únosnosti střechy pro novou rekreační oblast" strukturované a polostrukturované údaje - výpočtové tabulky).
- ⦿ **Vedoucí nákupu:** "Potřebujete specifikace a množství materiálů pro rekreační oblast, abyste mohli zorganizovat nákup" (textová a polostrukturovaná data - seznamy a specifikace).
- ⦿ **Inženýr pro bezpečnost a ochranu zdraví:** "Obdržel jsem údaje o nové oblasti. Vyhodnocuji rizika a aktualizuji bezpečnostní plán" polostrukturované údaje - dokumenty a plány).
- ⦿ **Specialista na BIM - modelování:** "Provádění změn v celkovém modelu projektu za účelem úpravy pracovní dokumentace" (geometrická data a polostrukturovaná data).
- ⦿ **Vedoucí projektu:** "Novou odpočívku začleňuji do harmonogramu prací. Aktualizuji harmonogramy a zdroje v systému řízení projektu" (strukturovaná a polostrukturovaná data - harmonogramy a plány).
- ⦿ **Specialista na údržbu zařízení (FM):** "Připravuji data pro budoucí údržbu rekreačního areálu a zadávám je do systému správy majetku" (strukturovaná a polostrukturovaná data - pokyny a plány údržby).



Obr. 3.1-3 Komunikace mezi specialisty probíhá na úrovni textu i dat.

Každý odborník pracuje s různými typy dat, aby zajistil efektivní týmovou spolupráci a úspěch projektu. Pochopení rozdílů mezi strukturovanými, polostrukturovanými a nestrukturovanými daty vám umožní rozpoznat jedinečnou roli, kterou každý typ hraje v digitálních podnikových procesech. Je důležité nejen vědět, že existují různé formy dat, ale také pochopit, jak, kde a proč se používají.

Není to tak dávno, co se myšlenka kombinovat tak různorodá data zdála být ambiciozní, ale obtížně realizovatelná. Dnes je již součástí každodenní praxe. Integrace dat různých schémat a struktur se stala nedílnou součástí architektury moderních informačních systémů.

V následujících kapitolách se budeme podrobně zabývat klíčovými standardy a přístupy, které umožňují spojit strukturovaná, částečně strukturovaná a nestrukturovaná data do jednoho uceleného pohledu. Zvláštní pozornost bude věnována strukturovaným datům a relačním databázím jako hlavním mechanismům pro ukládání, zpracování a analýzu informací ve stavebnictví.

Strukturovaná data

Ve stavebnictví pocházejí informace z mnoha zdrojů - výkresů, specifikací, plánů a zpráv. Aby bylo možné tento tok informací efektivně řídit, je třeba je strukturovat. Strukturovaná data umožňují uspořádat informace do pohodlné, čitelné a dostupné podoby.

Podle 5. výroční zprávy o stavebních technologiích společnosti JB Knowledge [17] 67% odborníků na řízení stavebních projektů sleduje a vyhodnocuje pracovní výkon ručně nebo pomocí tabulek.

Mezi nejběžnější formáty strukturovaných dat patří XLSX a CSV. Jsou široce používány pro ukládání, zpracování a analýzu informací v tabulkových procesorech. V těchto tabulkách jsou data prezentována ve formě řádků a sloupců, což usnadňuje jejich čtení, úpravu a analýzu.

XLSX, formát vytvořený společností Microsoft, je založen na použití XML -struktur a je archivován pomocí algoritmu ZIP. Hlavní rysy tohoto formátu:

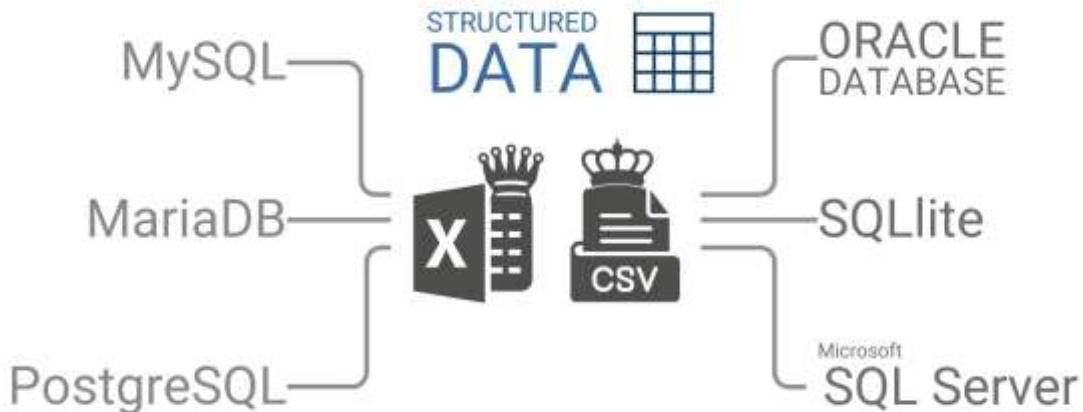
- Podpora složitých vzorců, grafů a maker.
- Možnost ukládat data do různých listů a formátovat informace.
- Optimalizováno pro Microsoft Excel, ale je kompatibilní i s jinými kancelářskými balíky.

Formát CSV je prostý textový soubor, ve kterém jsou hodnoty odděleny čárkami, středníky nebo jinými oddělovacími znaky. Hlavní výhody:

- Univerzální kompatibilita s různými programy a operačními systémy.
- Snadný import/export do databází a analytických systémů.
- Snadné zpracování i v textových editorech.

CSV však nepodporuje vzorce a formátování, takže jeho hlavním využitím je výměna dat mezi systémy a hromadná aktualizace informací. Díky své univerzálnosti a nezávislosti na platformě se CSV stal oblíbeným nástrojem pro přenos dat v heterogenních IT prostředích.

Dva formáty XLSX a CSV slouží jako spojovací článek mezi různými systémy pracujícími se strukturovanými daty (obrázek 3.1-4). Jsou užitečné zejména v úlohách, kde je důležitá čitelnost, ruční editace a základní kompatibilita.



Obr. 3.1-4 Formáty XLSX a CSV jsou spojovacím článkem mezi různými systémy, které pracují se strukturovanými daty.

Nezávislost na platformě činí z formátu CSV nejoblíbenější formát pro přenos dat v heterogenních IT prostředích a systémech.

XLSX a CSV však nejsou určeny pro vysoce výkonné výpočty nebo dlouhodobé ukládání velkého množství dat. Pro takové účely se používají modernější strukturované formáty, jako jsou Apache Parquet, Apache ORC, Feather, HDF5. Tyto formáty budou podrobněji rozebrány v kapitole "Ukládání velkých dat: Analýza populárních formátů a jejich efektivita" v 9. části této knihy.

V praxi se Excel s formátem XLSX častěji používá pro drobné úlohy a automatizaci rutinních procesů. Složitější scénáře vyžadují použití systémů pro správu dat, jako jsou ERP, PMIS CAFM, CPM, SCM a další (obr. 3.2-1). Tyto systémy uchovávají strukturovaná data, na nichž je založena organizace a řízení informačních toků v podniku.

Moderní informační systémy pro správu dat používané ve stavebnictví se opírají o strukturovaná data uspořádaná do tabulek. Pro spolehlivou, škálovatelnou a ucelenou správu velkých objemů informací se vývojáři aplikací a systémů obracejí k relačním systémům správy databází (RDBMS).

Relační databáze RDBMS a dotazovací jazyk SQL

K efektivnímu ukládání, zpracování a analýze dat slouží **relační databáze (RDBMS)**, což jsou systémy pro ukládání dat, které organizují informace do tabulek s definovanými vztahy mezi nimi.

Data uspořádaná v databázích (RDBMS) nejsou jen digitálními informacemi; jsou základem pro transakce a interakce mezi různými systémy.

Zde je několik nejběžnějších relačních systémů pro správu databází (RDBMS) (obrázek 3.1-5):

- **MySQL** (Open Source) je jedním z nejpopulárnějších RDBMS, který je součástí LAMP stacku (Linux, Apache, MySQL, PHP /Perl/Python). Díky své jednoduchosti a vysokému výkonu je hojně využíván při vývoji webových stránek.
- **PostgreSQL** (Open Source) je výkonný objektově-relační systém známý svou spolehlivostí a pokročilými funkcemi. Je vhodný pro komplexní podniková řešení.
- **Microsoft SQL Server** je komerční systém od společnosti Microsoft, který je široce používán ve firemním prostředí díky své integraci s dalšími produkty společnosti a vysoké úrovni zabezpečení.
- **Databáze Oracle** je jedním z nejvýkonnějších a nejspolehlivějších DBMS používaných ve velkých podnicích a kritických aplikacích.
- **IBM DB2** - zaměřený na velké společnosti, poskytující vysoký výkon a odolnost proti chybám.
- **SQLite** (Open Source) je lehká vestavěná databáze, ideální pro mobilní aplikace a samostatné systémy, jako je návrhový software CAD (BIM).

Oblíbené systémy pro správu databází ve stavebnictví - MySQL, PostgreSQL, Microsoft SQL Server, Oracle® Database, IBM® DB2 a SQLite - pracují se strukturovanými daty. Všechny tyto DBMS představují výkonná a flexibilní řešení pro správu široké škály podnikových procesů a aplikací, od malých webových stránek až po rozsáhlé podnikové systémy (obr. 3.2-1).

Podle Statista [48] budou relační systémy pro správu databází (RDBMS) v roce 2022 představovat přibližně 72% všech používaných DBMS.

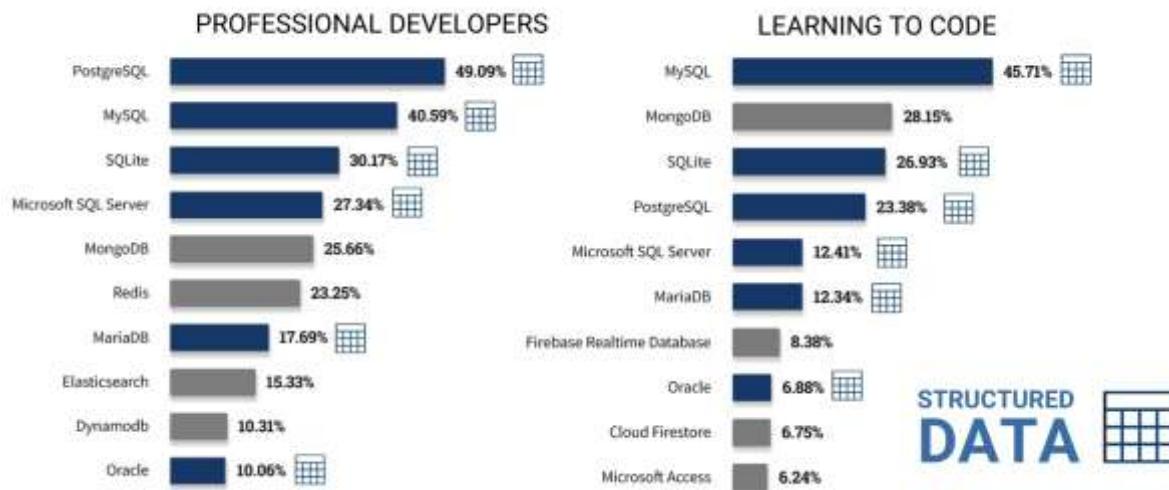
	Rank			DBMS	Database Model	Open Source vs Commercial
	Mar2025	Feb2025	Mar2024			
	1.	1.	1.	Oracle®	Relational, Multi-model	Commercial
	2.	2.	2.	MySQL	Relational, Multi-model	Open Source
	3.	3.	3.	Microsoft® SQL Server	Relational, Multi-model	Commercial
	4.	4.	4.	PostgreSQL	Relational, Multi-model	Open Source
	5.	5.	5.	MongoDB	Document, Multi-model	Open Source
	6.	7.	9.	Snowflake®	Relational	Commercial
	7.	6.	6.	Redis®	Key-value, Multi-model	Open Source
	8.	8.	7.	Elasticsearch®	Multi-model	Open Source
	9.	9.	8.	IBM Db2	Relational, Multi-model	Commercial
	10.	10.	10.	SQLite	Relational	Open Source
	11.	11.	12.	Apache Cassandra®	Multi-model	Open Source
	12.	12.	11.	Microsoft Access®	Relational	Open Source
	13.	13.	17.	Databricks®	Multi-model	Commercial
	14.	14.	13.	MariaDB	Relational, Multi-model	Open Source
	15.	15.	14.	Splunk	Search engine	Commercial
	16.	16.	16.	Amazon DynamoDB	Multi-model	Commercial
	17.	17.	15.	Microsoft Azure SQL	Relational, Multi-model	Commercial

Obr. 3.1-5 Oblíbenost používání strukturovaných databází (vyznačeno modře) v žebříčku DBMS (podle [49]).

Instalace databází s otevřeným zdrojovým kódem je poměrně snadná i bez rozsáhlých technických znalostí. Systémy s otevřeným zdrojovým kódem, jako je PostgreSQL, MySQL nebo SQLite, jsou k dispozici zdarma a fungují na většině operačních systémů: Windows, MacOS a Linux. Stačí přejít na oficiální webové stránky projektu, stáhnout si instalaci programu a postupovat podle pokynů. Ve většině případů netrvá instalace déle než 10-15 minut. Ve čtvrté části knihy si jednu takovou databázi vymodelujeme a vytvoříme (obr. 4.3-8).

Pokud vaše společnost využívá cloudové služby (například Amazon Web Services, Google Cloud nebo Microsoft Azure), můžete databázi nasadit na několik kliknutí - platforma vám nabídne připravené šablony pro instalaci. Díky otevřenosti kódu lze takové databáze snadno přizpůsobit vašim úkolům a obrovská komunita uživatelů vám vždy najít řešení jakéhokoli problému.

RDBMS zůstávají základem pro řadu podnikových aplikací a analytických platform (obrázek 3.1-6), které firmám umožňují efektivně ukládat, zpracovávat a analyzovat data - a tedy přijímat informovaná a včasná rozhodnutí.



Obr. 3.1-6 Průzkum mezi vývojáři na StackOverflow (největší IT fórum) o tom, které databáze používali v loňském roce a které chtějí používat v roce příštím (RDBMS jsou zvýrazněny modře) (podle [50]).

RDBMS poskytují spolehlivost, konzistenci dat, podporu transakcí a používají výkonné dotazovací jazyk SQL (Structured Query Language), který se často používá v analytice a umožňuje snadno získávat, upravovat a analyzovat informace uložené v databázích. SQL je hlavním nástrojem pro práci s daty v relačních systémech.

SQL - dotazy v databázích a nové trendy

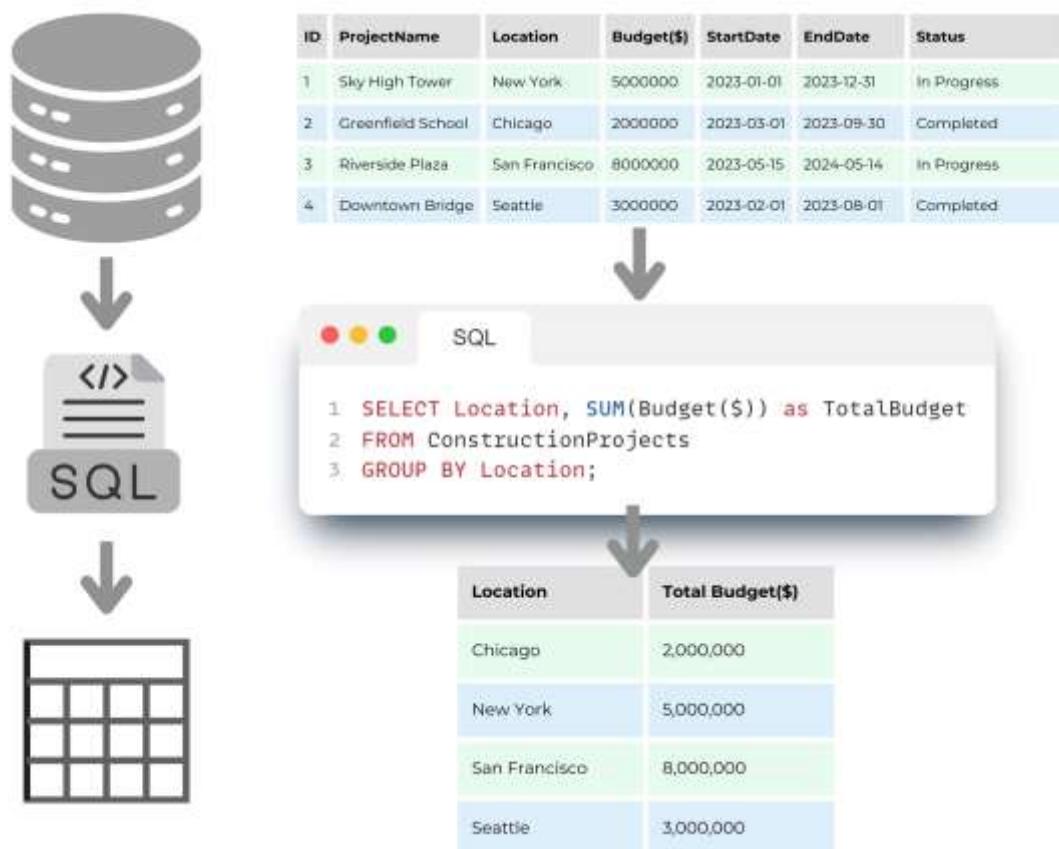
Hlavní výhodou jazyka SQL, často používaného v relačních databázích, oproti jiným typům správy informací (například pomocí klasických tabulek Excelu) je podpora velmi velkých objemů databází při vysoké rychlosti zpracování dotazů.

Strukturovaný dotazovací jazyk (SQL) je specializovaný programovací jazyk určený pro ukládání, zpracování a analýzu informací v relačních databázích. Jazyk SQL se používá k vytváření, správě a přístupu k datům a umožňuje efektivně vyhledávat, filtrovat, kombinovat a agregovat informace. Slouží jako klíčový nástroj pro přístup k datům a poskytuje pohodlný a formalizovaný způsob interakce s úložištěm informací.

Vývoj systémů SEQUEL-SQL prochází významnými produkty a společnostmi, jako jsou Oracle, IBM DB2, Microsoft SQL Server, SAP, PostgreSQL a MySQL, a vrcholí vznikem SQLite a MariaDB [51]. SQL poskytuje možnosti tabulkového procesoru, které v Excelu nenajdete, takže manipulace s daty je škálovatelnější, bezpečnější a snadno automatizovatelná:

- **Vytváření a správa datových struktur (DDL):** V SQL můžete vytvářet, upravovat a mazat tabulky v databázi, vytvářet vazby mezi nimi a definovat struktury pro ukládání dat. Excel naproti tomu pracuje s pevně danými listy a buňkami bez jasně definovaných vztahů mezi listy a datovými sadami.

- **Manipulace s daty (DML):** SQL umožňuje hromadně přidávat, upravovat, mazat a načítat data vysokou rychlostí pomocí složitých dotazů s filtrováním, tříděním a spojováním tabulek (obrázek 3.1-7). V aplikaci Excel vyžaduje zpracování velkého množství informací ruční úkony nebo speciální makra, což zpomaluje proces a zvyšuje pravděpodobnost chyb.
- **Řízení přístupu (DCL):** SQL umožňuje rozlišit přístupová práva k datům pro různé uživatele a omezit možnost upravovat nebo zobrazovat informace. Naproti tomu v aplikaci Excel je přístup buď sdílený (při přenosu souboru), nebo vyžaduje složité nastavení se sdílením oprávnění prostřednictvím clouдовých služeb.



Obr. 3.1-7 Příklad DML v SQL: rychlé zpracování, seskupování a agregace pomocí několika řádků kódu pro automatické zpracování dat.

Excel usnadňuje práci s daty díky své vizuální a intuitivní struktuře. S rostoucím množstvím dat však výkon Excelu klesá. Excel se také potýká s omezením množství dat, které může uložit - maximálně jeden milion řádků - a výkon se snižuje dlouho před dosažením tohoto limitu. Zatímco tedy Excel vypadá vhodněji pro vizualizaci a manipulaci s malým množstvím dat, SQL je vhodnější pro práci s velkými soubory dat.

Další etapou ve vývoji strukturovaných dat byl vznik sloupcových databází (Columnar Databases), které jsou alternativou k tradičním relačním databázím, zejména pokud jde o výrazně větší objemy dat a analytické výpočty. Na rozdíl od řádkových databází, kde se data ukládají po řádcích, sloupcové databáze zaznamenávají informace po sloupcích. Oproti klasickým databázím to umožňuje:

- Snížení úložného prostoru účinnou kompresí jednotných dat ve sloupcích.
- Zrychlení analytických dotazů, protože se načítají pouze požadované sloupce, nikoli celá tabulka.
- Optimalizace velkých dat a datových skladů, např. Data Lakehouse Architecture.

O sloupcových databázích, Pandas DataFrame, Apache Parquet, HDF5, stejně o vytváření velkých datových -skladů na jejich základě pro účely analýzy a zpracování dat si povíme více v následujících kapitolách této knihy - "DataFrame: univerzální formát tabulkových dat" a "Formáty ukládání dat a práce s Apache Parquet: DWH -datové sklady a architektura Data Lakehouse".

Nestrukturovaná data

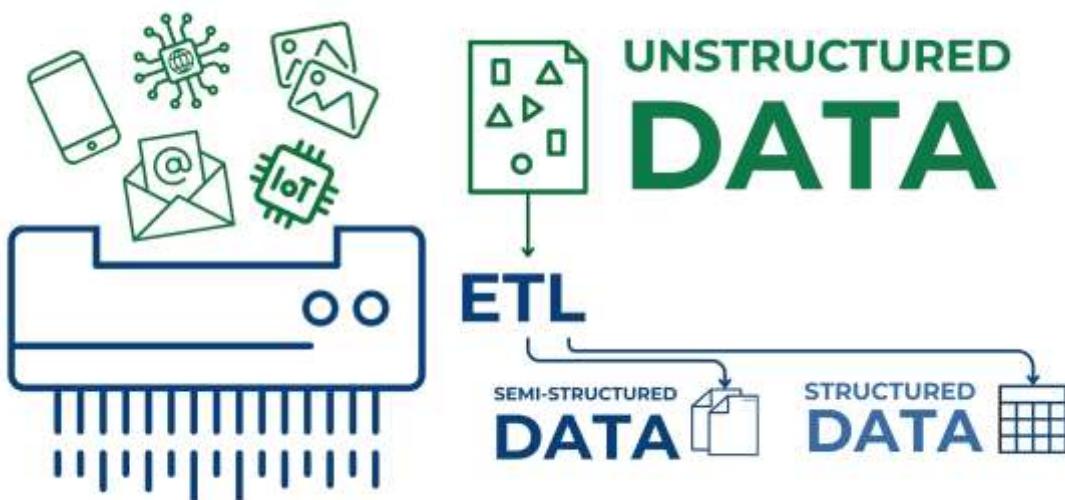
Ačkoli většina dat používaných v aplikacích a informačních systémech má strukturovanou podobu, většina informací generovaných ve stavebnictví má podobu nestrukturovaných dat - obrázků, videí, textových dokumentů, zvukových záznamů a dalších forem obsahu. To platí zejména ve fázích výstavby, provozu a technického dozoru, kde převažují vizuální a textové informace.

Nestrukturovaná data jsou informace, které nemají předem definovaný model nebo strukturu, nejsou uspořádány do tradičních řádků a sloupců jako v databázích nebo tabulkách.

Obecně lze nestrukturovaná data rozdělit do dvou kategorií:

- Nestrukturovaná data vytvořená člověkem, která zahrnují různé typy obsahu vytvořeného člověkem: textové dokumenty, e-maily, obrázky, videa atd.
- Strojově generovaná nestrukturovaná data jsou vytvářena zařízeními a senzory: patří mezi ně například soubory protokolů, data GPS, výsledky internetu věcí (IoT) a další telemetrické informace ze staveniště.

Na rozdíl od strukturovaných dat, která lze pohodlně uspořádat do tabulek a databází, vyžadují nestrukturovaná data před začleněním do informačních systémů další kroky zpracování (obr. 3.1-8). Využití technologií pro automatizovaný sběr, analýzu a transformaci takových dat otevří nové možnosti, jak zvýšit efektivitu výstavby, snížit chybovost a minimalizovat vliv lidského faktoru.



Obrázek 3.1-8 Zpracování nestrukturovaných dat začíná jejich transformací na polostrukturovaná a strukturovaná data.

Nestrukturovaná data tvoří až 80% všech informací [52], se kterými se odborníci ve firmách setkávají, proto se jejich typy a zpracováním budeme podrobně zabývat na příkladech v následujících kapitolách knihy.

Pro usnadnění diskuse jsou textové údaje rozděleny do samostatných kategorií. Přestože se jedná o typ spíše nestrukturovaných dat, jejich význam a rozšířenost ve stavebnictví vyžadují zvláštní pozornost.

Textová data: mezi nestrukturovaným chaosem a strukturovaným chaosem y

Textová data ve stavebnictví zahrnují širokou škálu formátů a typů informací, od papírových dokumentů až po neformální způsoby komunikace, jako jsou dopisy, rozhovory, pracovní korespondence a ústní jednání na staveništi. Všechna tato textová data nesou důležité informace pro řízení stavebních projektů, od podrobností o rozhodnutích o návrhu a změnách v plánech až po diskuse o bezpečnostních otázkách a jednání s dodavateli a zákazníky (obrázek 3.1-9).



Obr. 3.1-9 Textová data, jeden z nejoblíbenějších typů informací používaných při komunikaci mezi účastníky projektu.

Textové informace mohou být formalizované i nestrukturované. Mezi formalizované údaje patří dokumenty ve formátu Word (.doc,.docx), PDF, jakož i textové soubory zápisů z jednání (.txt). Neformalizovaná data zahrnují korespondenci messengerů a e-mailů, zápisy ze schůzek (Teams, Zoom, Google Meet) a zvukové záznamy diskusí (.mp3,.wav), které vyžadují převod do textové podoby.

Zatímco písemné dokumenty, jako jsou formální žádosti, smluvní podmínky a e-maily, již obvykle mají určitou strukturu, ústní komunikace a pracovní korespondence často zůstávají nestrukturované, což ztěžuje jejich analýzu a integraci do systémů řízení projektů.

Klíčem k efektivní správě textových dat je jejich převod do strukturovaného formátu. To umožňuje automatickou integraci zpracovaných informací do stávajících systémů, které již pracují se strukturovanými daty.



Obrázek 3.1-10 Převod textového obsahu na strukturovaná data.

Aby bylo možné efektivně využívat textové informace, je třeba je automaticky převést do strukturované podoby (obrázek 3.1-10). Tento proces obvykle zahrnuje několik kroků:

- **Rozpoznávání textu (OCR)** - převod obrázků dokumentů a výkresů do strojově čitelného formátu.
- **Analýza textu (NLP)** - automatická identifikace klíčových parametrů (data, částky a čísla relevantní pro projekt).
- **Klasifikace dat** - kategorizace informací (finance, logistika, řízení rizik).

Po rozpoznání a klasifikaci lze již strukturovaná data integrovat do databází a využívat je v automatizovaných systémech výkaznictví a řízení.

Polostrukturovaná a volně strukturovaná data

Polostrukturovaná data obsahují určitou úroveň uspořádání, ale nemají striktní schéma nebo strukturu. Ačkoli takové informace obsahují strukturované prvky (např. data, jména zaměstnanců a seznamy splněných úkolů), formát prezentace se může značně lišit projekt od projektu nebo dokonce i u jednotlivých zaměstnanců. Příkladem takových údajů jsou časové deníky, zprávy o postupu prací a harmonogramy, které mohou být prezentovány v různých formátech.

Polostrukturovaná data se analyzují snadněji než nestrukturovaná data, ale vyžadují další zpracování

pro integraci do standardizovaných systémů řízení projektů.

Práce s polostrukturovanými daty, jejichž struktura se neustále mění, představuje značnou výzvu. Proměnlivost struktury dat totiž vyžaduje samostatné individuální přístupy ke zpracování a analýze každého zdroje polostrukturovaných dat.

Zatímco práce s nestrukturovanými daty vyžaduje velké úsilí, zpracování polostrukturovaných dat lze provádět pomocí relativně jednoduchých metod a nástrojů.

Slabě strukturovaná data je obecnější termín, který popisuje data s minimální nebo neúplnou strukturou. Nejčastěji se jedná o textové dokumenty, chaty, e-mailsy, kde se vyskytují některá metadata (např. datum, odesíatel), ale většina informací je prezentována chaoticky.

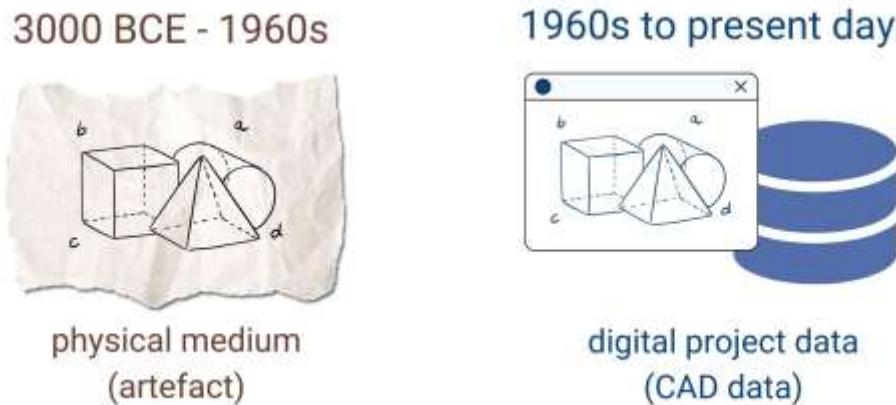
Ve stavebnictví se volně strukturovaná data vyskytují v různých procesech. Mohou zahrnovat například:

- Odhady a cenové nabídky - tabulky s údaji o materiálech, množství a nákladech, ale bez jednotného formátu.
- Výkresy a technická schémata - soubory ve formátu PDF nebo DWG, které obsahují textové poznámky a metadata, ale nemají strikně stanovenou strukturu.
- Pracovní harmonogramy - data z MS Project, Primavera P6 nebo jiných systémů, které mohou mít odlišnou strukturu exportu.
- CAD (BIM -modely) - obsahují prvky konstrukce, ale reprezentace dat závisí na softwaru a standardu projektu.

Geometrická data, vytvářená systémy CAD, lze kategorizovat stejným způsobem jako polostrukturovaná data. Geometrická data CAD (BIM) však budeme klasifikovat jako samostatný typ dat, protože s nimi, podobně jako s textovými daty, lze ve firemních procesech často zacházet jako se samostatným typem dat.

Geometrická data a jejich použití

Zatímco metadata o prvcích projektu jsou téměř vždy uložena ve formě tabulek, strukturovaných nebo volně strukturovaných formátů, geometrická data prvků projektu na jsou ve většině případů vytvářena pomocí speciálních nástrojů CAD (obr. 3.1-11), které umožňují detailně vizualizovat prvky projektu jako soubor linií (2D) nebo geometrických těles (3D).

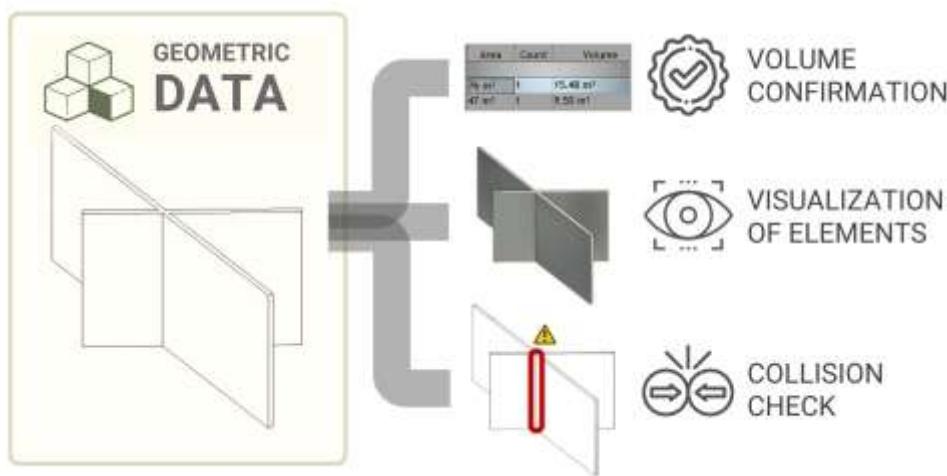


Obr. 3.1-11 Nástroje CAD pomohly přenést geometrické informace z fyzických médií do databázové podoby.

Při práci s geometrickými daty ve stavebnictví a architektuře lze rozlišit tři hlavní způsoby využití geometrických dat (obrázek 3.1-12):

- **Potvrzení objemu:** geometrické údaje, generované v programech CAD (BIM) pomocí speciálních geometrických jader, jsou potřebné k automatickému a přesnému určení objemů a rozměrů prvků projektu. Tato data zahrnují automaticky vypočítané plochy, objemy, délky a další důležité atributy potřebné pro plánování, rozpočtování a objednávání zdrojů a materiálů
- **Vizualizace projektu:** v případě jakýchkoli změn v projektu umožňuje vizualizace prvků automatické generování aktualizovaných výkresů v různých rovinách. Vizualizace projektu v počátečních fázích umožňuje rychlejší porozumění mezi všemi účastníky a šetří tak čas a zdroje v průběhu výstavby.
- **Kontrola kolizí:** V komplexních stavebních a inženýrských projektech, kde je rozhodující interakce více kategorií prvků (např. potrubí a stěn) bez "geometrických konfliktů", hraje kontrola kolizí klíčovou roli. Použití softwaru pro zjišťování kolizí umožňuje proaktivně identifikovat potenciální geometrické konflikty mezi prvky projektu a předcházet tak nákladným chybám v průběhu výstavby.

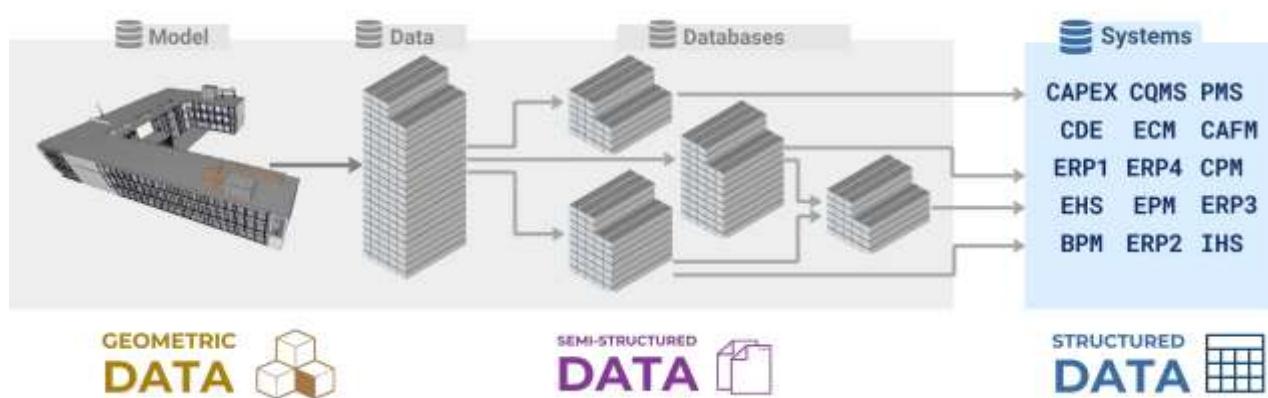
Od samého počátku existence inženýrských projekčních kanceláří, od doby výstavby prvních složitých staveb, poskytovali stavební inženýři geometrické informace ve formě výkresů, čar a plošných geometrických prvků (na papyru, papíru "A0" nebo ve formátech DWG, PDF, PLT), na jejichž základě mistři a odkadci (obr. 3.1-11), v průběhu posledních tisíciletí pomocí pravítek a úhloměrů shromažďovali atributové objemy nebo počty prvků a skupin prvků (3.1-11), za poslední tisíciletí pomocí pravítek a transportérů shromažďovali atributové objemy neboli množství prvků a skupin prvků.



Obr. 3.1-12 Geometrie je základem pro získání objemových parametrů prvků, které se následně používají pro výpočet nákladů a časového plánu projektu.

Dnes je tento ruční a časově náročný úkol vyřešen plnou automatizací díky nástupu objemového modelování v moderních nástrojích CAD (BIM), které umožňuje automaticky, pomocí speciálního geometrického jádra, získat objemové atributy libovolného prvku bez nutnosti ručního výpočtu objemových parametrů.

Moderní nástroje CAD také umožňují klasifikovat a kategorizovat prvky projektu, takže můžete z databáze projektu nahrát tabulky specifikací pro použití v různých systémech, jako je odhad nákladů, plánování nebo výpočet CO₂ (obr. 3.1-13). Získávání specifikací, tabulek QTO a množství a praktickým příkladům se budeme věnovat v kapitole "Získávání množství a kvantifikace".



Obr. 3.1-13 Nástroje CAD (BIM) ukládají data do databází, které jsou navrženy tak, aby se integrovaly a spolupracovaly s dalšími systémy.

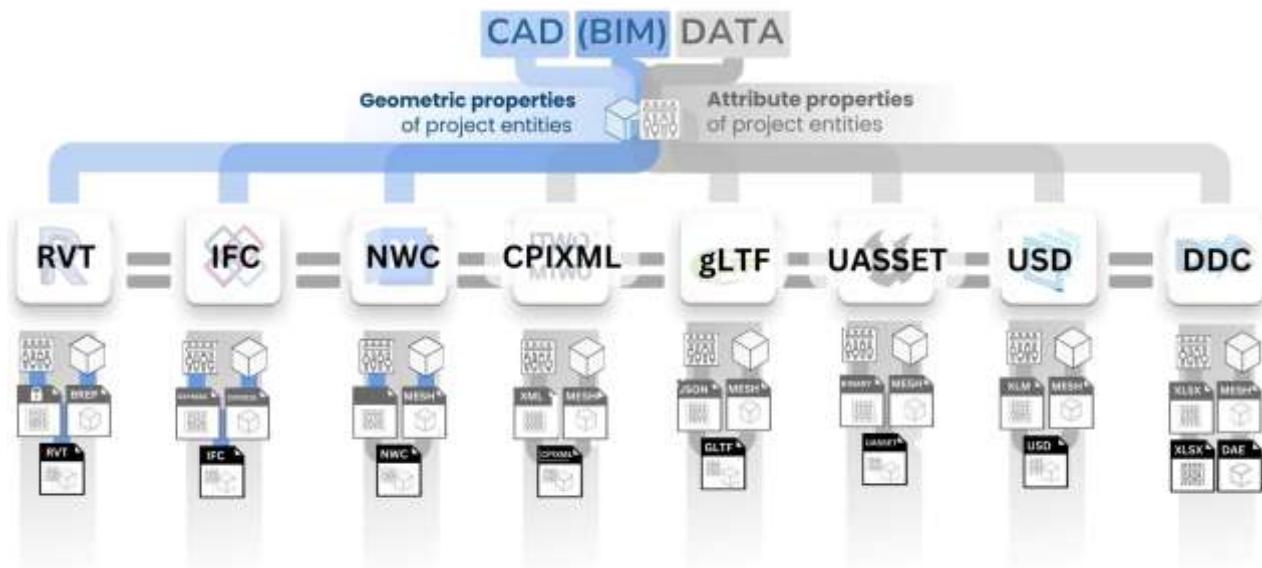
Vzhledem k uzavřené povaze databází a formátů používaných v prostředí CAD se geometrická data vytvořená v řešeních CAD stala vlastně samostatným typem informací. Spojují v sobě jak geometrii prvků, tak metainformace (strukturované nebo polostrukturované), uzavřené ve specializovaných souborech a formátech.

Data CAD: od návrhu po ukládání dat

Moderní systémy CAD a BIM ukládají data ve vlastních, často proprietárních formátech: DWG, DXF, RVT, DGN, PLN a dalších. Tyto formáty podporují 2D i 3D reprezentaci objektů a zachovávají nejen geometrii, ale také atributy spojené s objekty. Zde jsou uvedeny ty nejběžnější:

- **DWG** je binární formát souboru, který se používá k ukládání dvourozměrných (a méně často trojrozměrných) návrhových dat a metadat.
- **DXF** je textový formát pro výměnu 2D a 3D výkresů mezi systémy CAD. Obsahuje geometrii, vrstvy a atributová data, podporuje ASCII i binární reprezentaci.
- **RVT** je binární formát pro ukládání modelů CAD včetně 3D -geometrie, atributů prvků, vztahů a parametrů návrhu.
- **IFC** je otevřený textový formát pro výměnu stavebních dat mezi systémy CAD (BIM). Obsahuje geometrii, vlastnosti objektů a informace o jejich vzájemných vztazích.

Kromě těchto formátů se používají i další: PLN, DB1, SVF, NWC, CPIXML, BLEND, BX3, USD, XLSX, DAE. Ačkoliv se liší účelem a mírou otevřenosti (obrázek 3.1-14), všechny mohou reprezentovat stejný informační model projektu v různých formách. V komplexních projektech se tyto formáty často používají souběžně, od výkresů až po koordinační modely projektu.



Obrázek 3.1-14 Oblíbené formáty pro ukládání CAD popisují geometrii pomocí parametrů BREP nebo MESH doplněných atributovými údaji.

Všechny výše uvedené formáty umožňují ukládat data o jednotlivých prvcích stavebního projektu a všechny výše uvedené formáty obsahují dva klíčové typy dat:

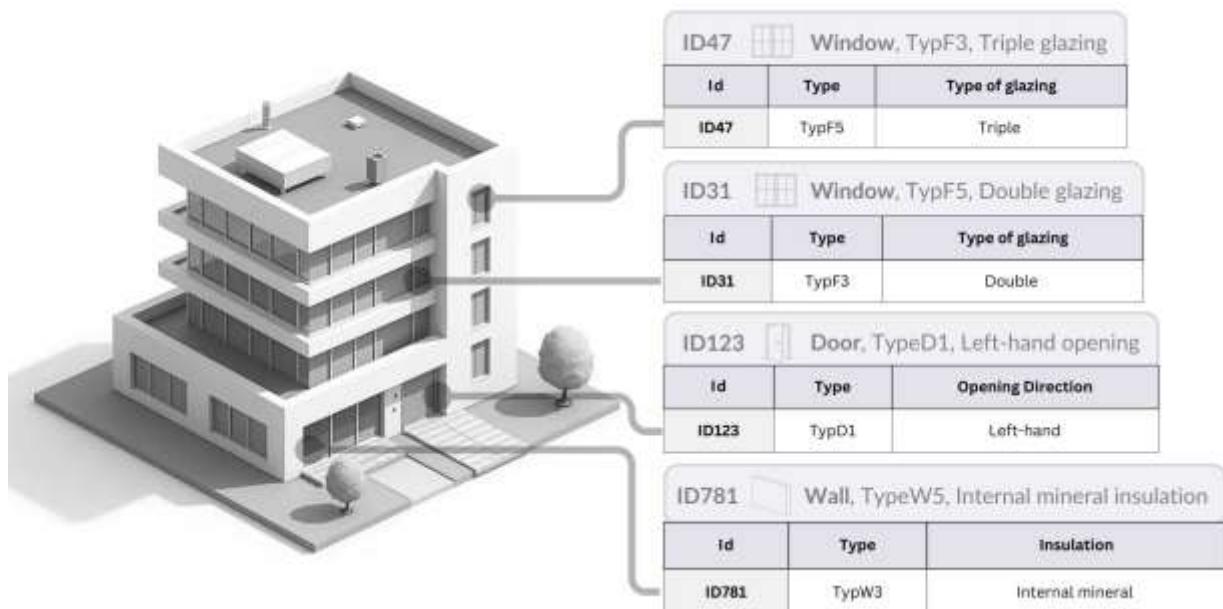
- **Geometrické parametry** - popisují tvar, umístění a rozměry objektu. Geometrii a jejímu využití se budeme podrobně věnovat v šesté části knihy věnované řešením CAD (BIM);
- **Vlastnosti atributů** - obsahují různé informace: materiály, typy prvků, technické charakteristiky, jedinečné identifikátory a další vlastnosti, které mohou mít prvky projektu.

Atributové údaje mají v moderních projektech zvláštní význam, protože definují provozní vlastnosti objektů, umožňují inženýrské výpočty a kalkulace nákladů a zajišťují komplexní interakci mezi účastníky projektování, výstavby a provozu. Např.:

- U oken a dveří: typ konstrukce, typ zasklení, směr otevírání (obr. 3.2-1).
- U stěn se zaznamenávají informace o materiálech, tepelné izolaci a akustických vlastnostech.
- U inženýrských sítí jsou uloženy parametry potrubí, kanálů, kabelových tras a jejich přípojek.

Tyto parametry mohou být uloženy jak v samotných souborech CAD-(BIM -), tak v externích databázích - jako výsledek exportu, konverze nebo přímého přístupu k interním strukturám CAD prostřednictvím nástrojů reverzního inženýrství. Tento přístup usnadňuje integraci informací o návrhu s dalšími podnikovými systémy a platformami

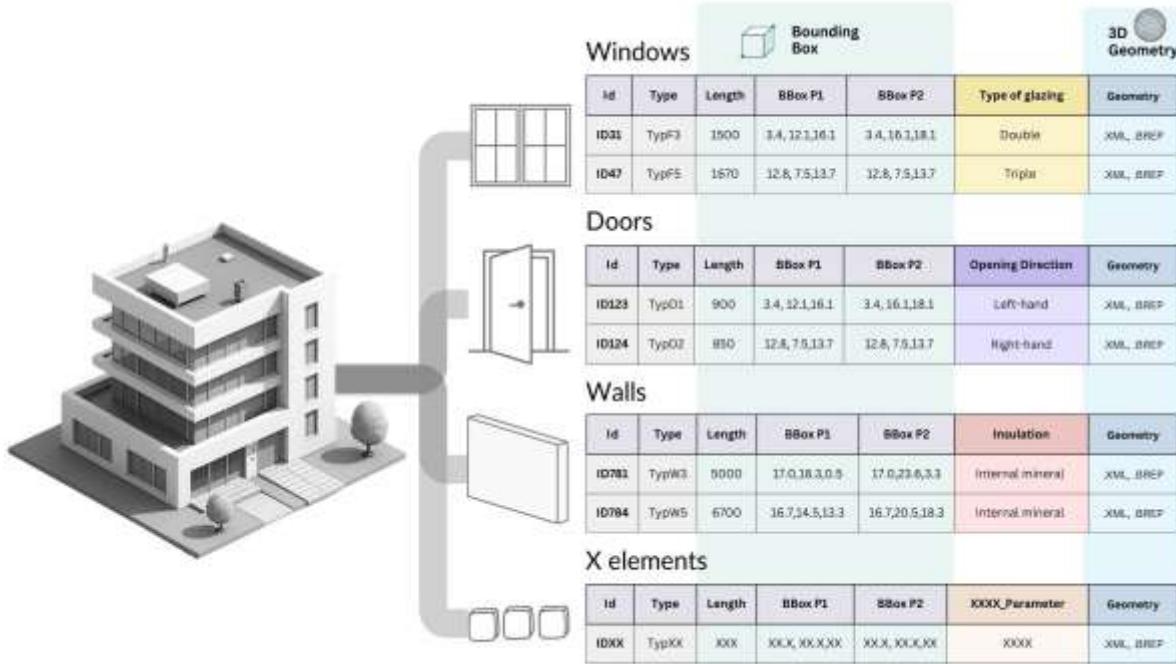
Reverzní inženýrství v kontextu CAD (BIM) je proces extrakce a analýzy vnitřní struktury digitálního modelu s cílem obnovit jeho logiku, strukturu dat a závislosti bez přístupu k původním algoritmu nebo dokumentaci.



Obr. 3.1-15 Prvek projektu kromě popisu parametrické nebo polygonální geometrie obsahuje informace o parametrech a vlastnostech prvků.

V důsledku toho se kolem každého prvku vytvoří jedinečná sada parametrů a vlastností, která zahrnuje jak jedinečné charakteristiky každého objektu (např. identifikátor a rozměry), tak společné atributy pro skupiny prvků. To umožňuje nejen analyzovat jednotlivé prvky-objekty projektu, ale také je spojovat do logických skupin, které pak mohou využívat další specialisté pro své úkoly a výpočty v systémech a databázích.

Entita je konkrétní nebo abstraktní objekt reálného světa, který lze jednoznačně identifikovat, popsat a reprezentovat ve formě dat.



Obr. 3.1-16 Každý prvek projektu obsahuje atributy, které bud' zadává projektant, nebo se vypočítávají v programu CAD.

V posledních desetiletích se ve stavebnictví vyvinulo mnoho nových formátů CAD (BIM), které zjednodušují tvorbu, ukládání a přenos dat. Tyto formáty mohou být uzavřené nebo otevřené, tabulkové, parametrické nebo grafické. Jejich rozmanitost a roztríštěnost však značně komplikuje správu dat ve všech fázích životního cyklu projektu. Srovnávací tabulka hlavních formátů používaných pro výměnu informací ve stavebnictví je uvedena na obrázku 3.1-17 (plná verze je k dispozici pomocí QR kódu).

Na řešení problémů interoperability a přístupu k datům CAD se podílejí správci BIM (BIM) a koordinátoři, jejichž úkolem je kontrolovat exporty, kontrolovat kvalitu dat a integrovat části dat CAD (BIM) do jiných systémů.

Vzhledem k uzavřené povaze a složitosti formátů je však obtížné tento proces automatizovat, což nutí odborníky provádět mnoho operací ručně, bez možnosti vytvořit plnohodnotné procesy in-line zpracování dat (pipeline).

The table is a complex data storage system for construction CAD (BIM) data. It includes sections for 'Geometric properties of project entities' and 'Attribute properties of project entities'. The data is organized into multiple columns representing different types of project entities and their properties. A QR code is provided for quick access to the data.

Obrázek 3.1-17 Tabulka porovnávající hlavní datové formáty, ve kterých jsou uloženy informace o prvcích projektu [53].

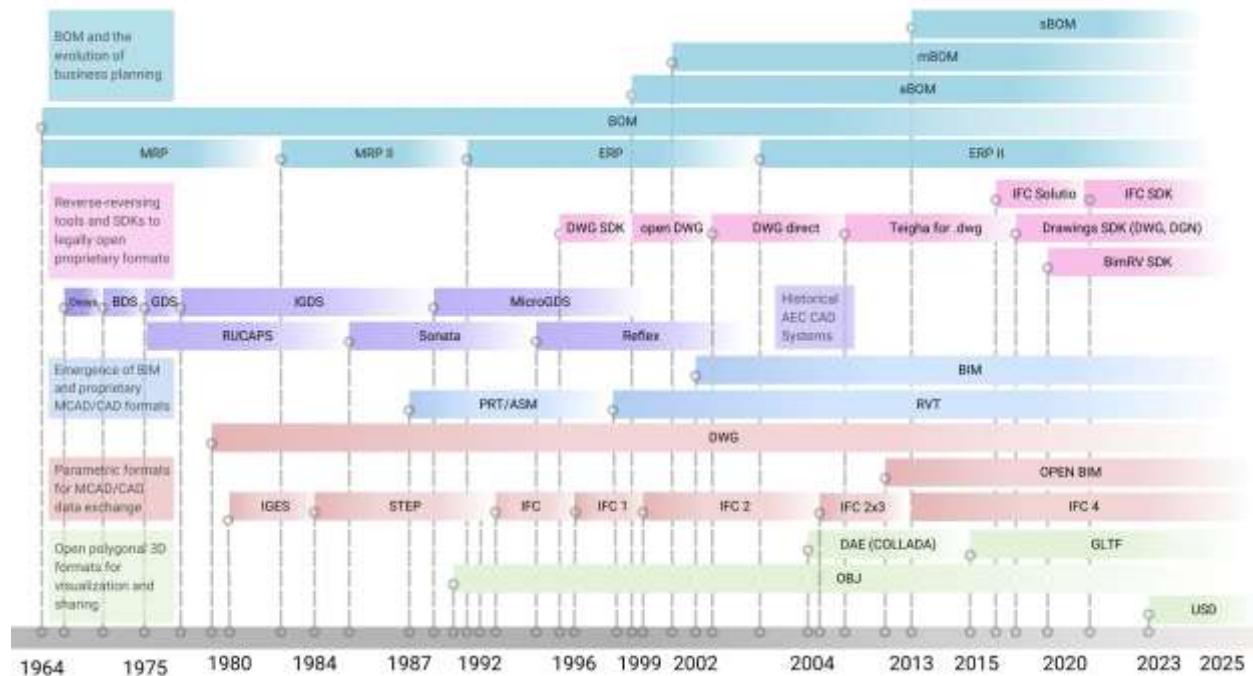
Abychom pochopili, proč existuje tak mnoho různých datových formátů a proč je většina z nich uzavřená, je důležité proniknout do procesů, které probíhají uvnitř programů CAD (BIM, kterým se budeme podrobně věnovat v šesté části knihy).

Další informační vrstvu přidanou ke geometrii zavedli vývojáři systémů CAD v podobě konceptu BIM (Building Information Modeling), což je marketingový termín aktivně prosazovaný ve stavebnictví od roku 2002 [54].

Vznik konceptu BIM (BOM) a využití CAD v procesech.

Koncepce informačního modelování budov (BIM), poprvé nastíněná v dokumentu BIM Whitepaper z roku 2002 [54], vznikla na základě marketingových iniciativ výrobců softwaru CAD. Vznikl z marketingových iniciativ výrobců CAD softwaru a byl pokusem přizpůsobit principy již dobře zavedené ve strojírenství potřebám stavebnictví.

Inspirací pro BIM byl koncept BOM (Bill of Materials) - kusovník, který se v průmyslu hojně používá od konce 80. let. Ve strojírenství umožnil BOM propojit data ze systémů CAD se systémy PDM (Product Data Management), PLM (Product Lifecycle Management) a ERP, a zajistit tak holistickou správu technických informací v průběhu celého životního cyklu výrobku (obr. 3.1-8).



Obr. 3.1-18 Vývoj specifikací (BOM), informačního modelování (BIM) a digitálních formátů v inženýrském stavebnictví.

Moderní vývoj konceptu BOM vedl ke vzniku rozšířeného rámce - XBOM (Extended BOM), který zahrnuje nejen složení výrobku, ale také scénáře chování, provozní požadavky, parametry udržitelnosti a údaje pro prediktivní analýzu. XBOM plní v podstatě stejnou úlohu jako BIM ve stavebnictví: oba přístupy se snaží proměnit digitální model v Jednotný zdroj pravdy pro všechny účastníky projektu v průběhu celého životního cyklu projektu.

Klíčovým milníkem v nástupu BOM ve stavebnictví bylo zavedení prvního parametrického CAD (MCAD) speciálně upraveného pro stavebnictví v roce 2002. Vyvinul jej tým, který předtím vytvořil Pro-E®, revoluční systém MCAD pro strojírenství, který vznikl koncem 80. let a stal se průmyslovým standardem [55].

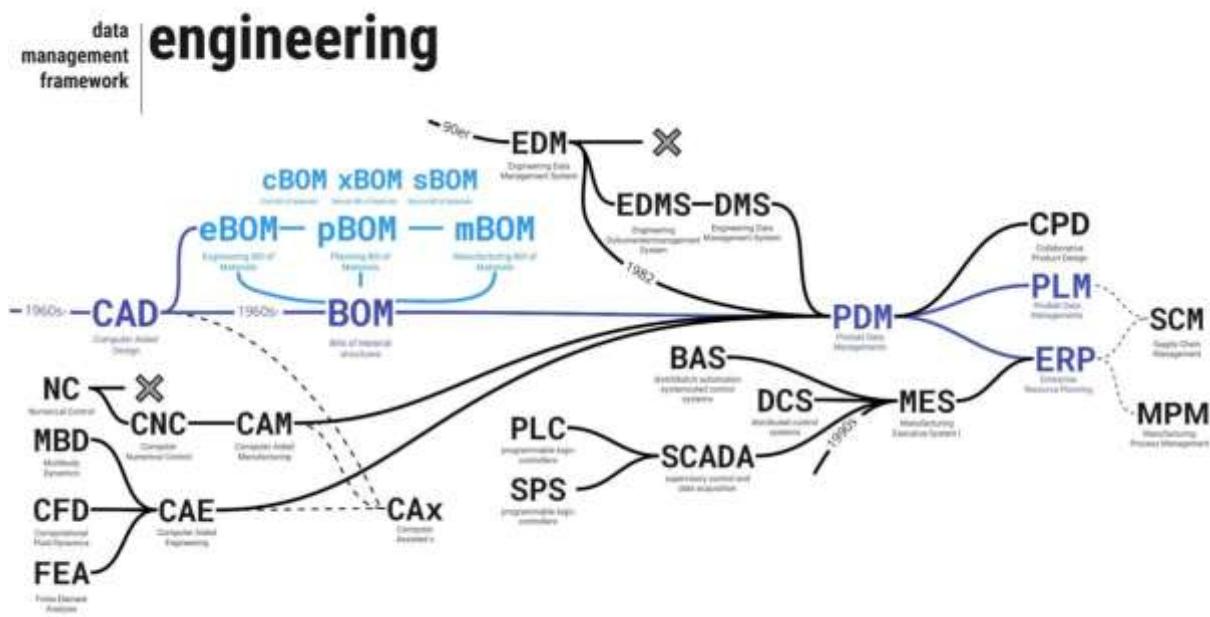
Již koncem 80. let bylo cílem odstranit omezení [56] tehdejších CAD -programů. Hlavním cílem bylo snížit pracnost při provádění změn parametrů konstrukčních prvků a umožnit aktualizaci modelu na základě údajů mimo CAD programy prostřednictvím databáze [57]. Nejdůležitější roli přitom měla hrát parametrizace: automatické načítání charakteristik z databáze a jejich využití k aktualizaci modelu uvnitř systémů CAD.

Pro-E a koncept elementárního parametrického modelování c BOM, který je jeho základem, měly významný vliv na vývoj trhu CAD - a MCAD [58]. Již 25 let je tento model v průmyslu a mnoha moderních systémů se stalo jeho koncepčními nástupci.

Cílem je vytvořit systém, který bude dostatečně flexibilní, aby inženýr mohl snadno zvažovat různé návrhy. A náklady na provedení změn návrhu by se měly co nejvíce blížit nule. Tradiční software CAD / CAM nerealisticky omezuje provádění levných změn pouze na samém začátku procesu návrhu [59].

- Samuel Heisenberg, zakladatel společnosti Parametric Technology Corporation®, vývojář MCAD -produkту Pro-E a učitel tvůrce CAD produktu využívajícího formát RVT

Ve strojírenství se systémy PDM, PLM, MRP a ERP staly klíčovými platformami. Hrají ústřední roli při správě dat a procesů, shromažďují informace z CAx systémů (CAD, CAM, CAE) a organizují konstrukční činnosti na základě struktury výrobku (BOM: eBOM, pBOM, mBOM) (obr. 3.1-18). Tato integrace snižuje chybovost, zabraňuje duplicitě dat a zajišťuje komplexní sledovatelnost od návrhu až po výrobu.



Obrázek 3.1-19 Historicky vznikl BOM v 60. letech 20. století jako způsob strukturování dat ze systémů CAx a jejich předávání řídicím systémům.

Nákup řešení CAD vyvinutého bývalým týmem Pro-E a založeného na přístupu BOM jedním z předních dodavatelů byl téměř okamžitě poznamenán zveřejněním série BIM Whitepaper (2002-2003)[60][61]. Již od poloviny roku 2000 se koncept BIM začal aktivně prosazovat ve stavebnictví, což výrazně zvýšilo zájem o parametrický software. Obliba rostla tak rychle, že stavební vidlice strojírenského softwaru Pro-E - parametrický CAD prosazovaný tímto výrobcem - skutečně vytlačil konkurenci v segmentu architektonického a konstrukčního navrhování (obr. 3.1-20). Počátkem roku 2020 si de facto upevnil globální dominanci na trhu BIM (CAD) [62].



Obr. 3.1-20 Popularita dotazů ve vyhledávači Google (RVT versus IFC): parametrický CAD vytvořený bývalým týmem Pro-E s podporou kusovníků -BIM získal popularitu téměř ve většině zemí světa.

Za posledních 20 let získala zkratka BIM mnoho výkladů, jejichž rozmanité významy mají kořeny v původních marketingových konцепcích, které vznikly na počátku roku 2000. Norma ISO 19650, která sehrála důležitou roli při popularizaci tohoto termínu, vlastně zajistila status BIM jako "vědecky podloženého" přístupu ke správě informací. V samotném textu normy, která se věnuje správě dat v celém životním cyklu objektů pomocí BIM, je však zkratka BIM zmíněna, ale nikdy není jasně definována

Původní webové stránky dodavatele, které v letech 2002 [60] a 2003 [61] zveřejnily řadu bílých knih o BIM, ve skutečnosti reprodukovaly marketingové materiály o konceptech BOM (Bills of Materials) a PLM (Product Lifecycle Management), které byly dříve používány ve strojírenském softwaru Pro-E již v 90. letech [63].

Informační modelování budov, inovativní nový přístup k navrhování, výstavbě a správě budov, který v roce 2002 představila společnost..... [název dodavatele CAD], změnil způsob, jakým odborníci z celého světa přemýšlejí o tom, jak lze technologii použít při navrhování, výstavbě a správě budov.

- BIM Whitepaper, 2003 [61].

Tyto první publikace přímo spojovaly BIM s konceptem centralizované integrované databáze. Jak se uvádí v bílé knize z roku 2003, BIM je správa informací o budově, kde všechny aktualizace probíhají v jediném úložišti, což zajišťuje synchronizaci všech výkresů, řezů a specifikací (BOM - Bills of Materials).

BIM je popisován jako správa informací o budově, kdy veškeré aktualizace a změny probíhají v databázi. Ať už se tedy jedná o schémata, řezy nebo výkresy listů, vše je vždy koordinované, konzistentní a aktuální.

- Webové stránky společnosti CAD vendor with BIM Whitepaper, 2003 [54].

Myšlenka řízení designu prostřednictvím jediné integrované databáze byla široce diskutována již ve výzkumu v 80. letech 20. století. Například koncepce BDS Charlese Eastmana [57] obsahovala 43 odkazů na termín "databáze" (obrázek 6.1-2). Do roku 2004 se tento počet snížil téměř o polovinu na 23 v Bílé knize o BIM z roku 2002 [64]. A do poloviny roku 2000 se téma databází prakticky vytratilo z marketingových materiálů dodavatelů a z agendy digitalizace obecně.

Přestože původně byla jádrem systému BIM databáze a přístup k ní, postupem času se důraz přesunul na geometrii, vizualizaci a 3D. Sám registrátor standardu IFC z roku 1994, který v roce 2002 vydal BIM Whitepaper - tentýž dodavatel - ve Whitepaper z počátku roku 2000 výslovně upozornil na omezení neutrálních formátů, jako jsou IGES, STEP a IFC, a na potřebu přímého přístupu k databázím CAD:

Různé aplikace mohou být nekompatibilní a znova zadané údaje mohou být nepřesné [...]. Výsledek tradičního počítačem podporovaného navrhování [CAD]: vyšší náklady, delší doba uvedení na trh a nižší kvalita výrobků. Dnes všechny hlavní aplikace používají pro výměnu dat na nízké úrovni standardní rozhraní. Použitím starých standardů IGES nebo nového STEP [IFC je de facto i de iure kopíí formátu STEP/IGES] pro výměnu dat mezi aplikacemi různých výrobců mohou uživatelé dosáhnout určité kompatibility dat mezi nejlepšími produkty. IGES a STEP však fungují pouze na nízkých úrovních a nemohou si vyměňovat tak bohatá data, jako jsou informace generované dnešními špičkovými aplikacemi [...]. A přestože se tyto a další standardy téměř denně zlepšují, vždy budou z hlediska bohatosti dat zaostávat za produkty dnešních výrobců. [...] Programy v rámci aplikace musí být schopny vyměňovat a zachovávat datové bohatství, aniž by se uchýlily k neutrálním překladačům, jako jsou IGES, STEP [IFC] nebo PATRAN. Místo toho by rámcové aplikace měly mít možnost přímého přístupu k podkladové databázi CAD, aby se neztratila podrobnost a přesnost informací.

- Bílá kniha dodavatele CAD (IFC, BIM) "Integrated Design and Manufacturing: Benefits and Rationale", 2000 [65].

Již v 80. letech a na počátku nového tisíciletí se tedy za klíčový prvek digitálního navrhování v prostředí CAD považovala databáze, nikoli formátový soubor nebo neutrální formát IFC. Navrhovalo se, aby se upustilo od překladačů a aplikace měly přímý přístup k datům. Ve skutečnosti však v polovině 20. let 20. století začal koncept BIM připomínat strategii "rozděl a panuj", kdy jsou zájmy výrobců softwaru používajícího uzavřená geometrická jádra upřednostňovány před rozvojem otevřené výměny informací.

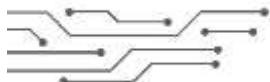
BIM je dnes vnímán jako nedílná součást stavebnictví. V uplynulých dvou desetiletích se však sliby zjednodušené spolupráce a integrace dat z velké části nenaplnily. Většina řešení je stále vázána na uzavřené nebo neutrální formáty a specializované nástroje. Historii BIM, otevřenému BIM a IFC, stejně jako problematice interoperability a geometrickým jádrům se budeme podrobně věnovat v 6. části knihy "CAD a BIM: Marketing, realita a budoucnost návrhových dat ve stavebnictví".

V současné době stojí odvětví před klíčovou výzvou přejít od tradičního chápání CAD (BIM) jako nástroje pro modelování k jeho využití jako plnohodnotné databáze. To vyžaduje nové přístupy k práci s informacemi, opuštění závislosti na uzavřených ekosystémech a zavedení otevřených řešení.

S rozvojem nástrojů reverzního inženýrství, které umožňují přístup k databázím CAD, a s rozšířením technologií Open Source a LLM se uživatelé a vývojáři ve stavebnictví stále více odklánějí od vágních podmínek dodavatelů softwaru. Místo toho se pozornost přesouvá na to, co je skutečně důležité: data (databáze) a procesy.

Za módními zkratkami a vizualizacemi se skrývají standardní postupy správy dat: ukládání, přenos a transformace - tj. klasický proces ETL (Extract, Transform, Load). Stejně jako v jiných odvětvích vyžaduje digitalizace stavebnictví nejen výměnné standardy, ale také jasně strukturované nakládání s heterogenními informacemi.

Aby bylo možné plně využít potenciál dat CAD (BIM), musí společnosti přehodnotit svůj přístup ke správě informací. To nevyhnutelně povede ke klíčovému prvku digitální transformace - sjednocení, standardizaci a smysluplnému strukturování dat, se kterými stavební odborníci denně pracují.

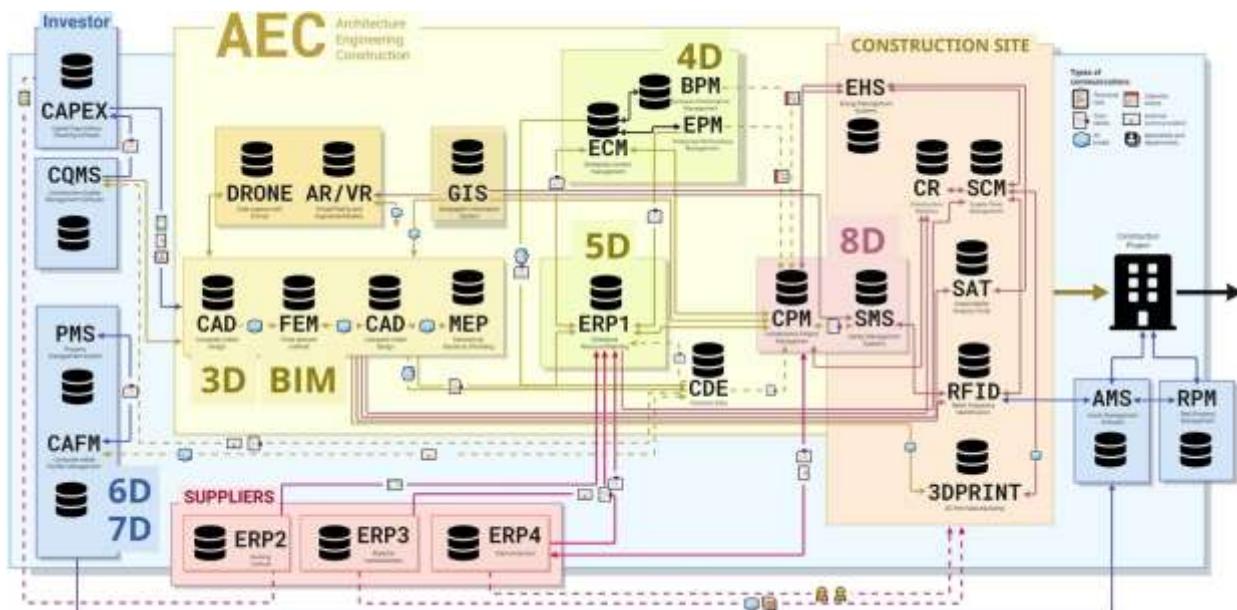


KAPITOLA 3.2.

HARMONIZACE A STRUKTUROVÁNÍ DAT

Naplňování systémů daty ve stavebnictví

Ať už se jedná o velké korporace nebo středně velké podniky, specialisté se denně zabývají naplňováním softwarových systémů a databází s různými rozhraními multiinformátovými informacemi (obr. 3.2-1), které musí s pomocí manažerů vzájemně spolupracovat. Právě tento komplex vzájemně se ovlivňujících systémů a procesů nakonec vytváří příjmy a zisk podniku.



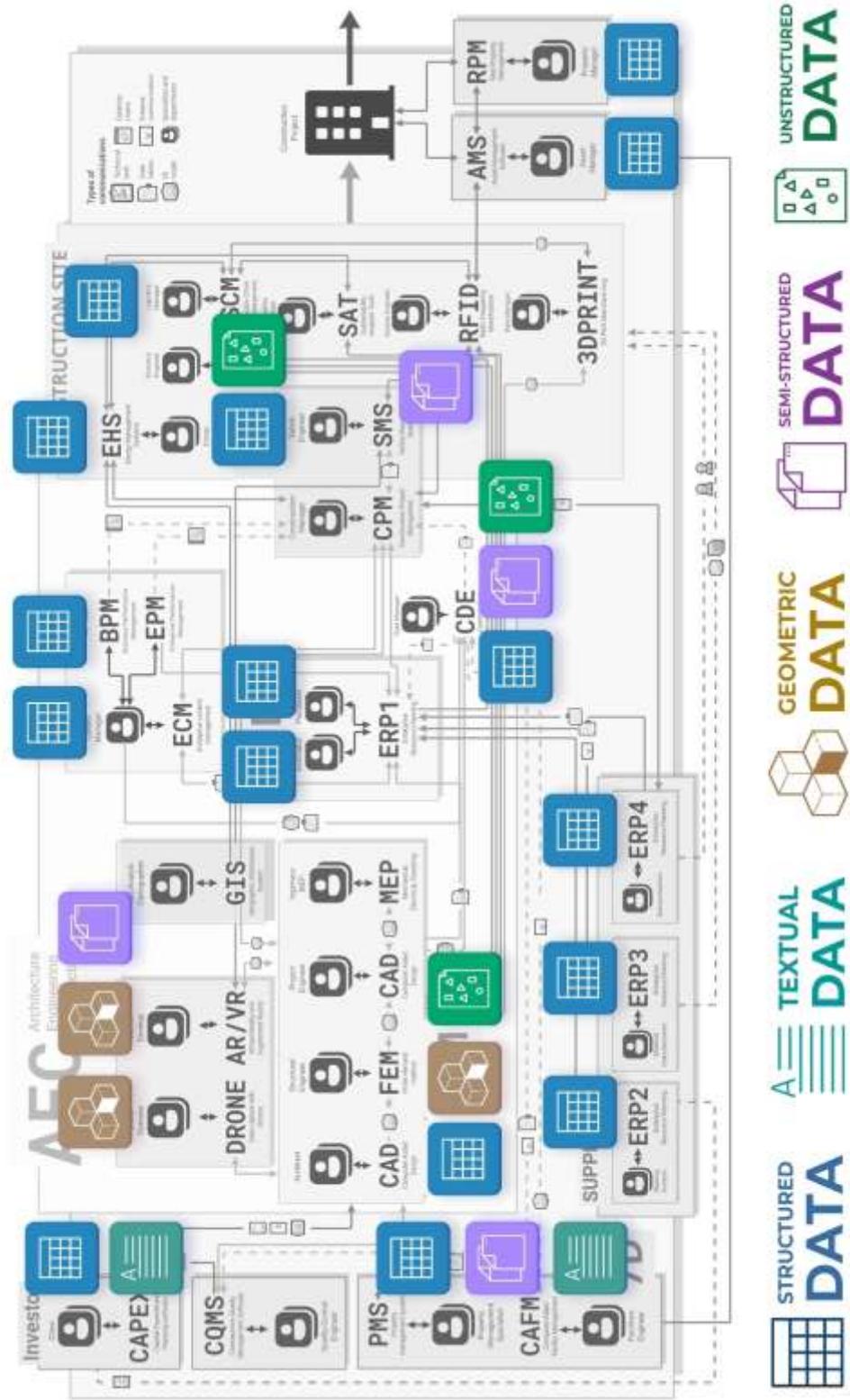
Obrázek 3.2-1 Prakticky každý systém nebo aplikace používaná ve stavebnictví má ve svém jádru některou z populárních databází RDBMS.

Každá z výše uvedených kategorií systémů, které se uplatňují ve stavebnictví, pracuje s vlastními datovými typy odpovídajícími funkční úloze těchto systémů. Abychom přešli od abstraktní úrovni ke konkrétní, přejdeme od datových typů k jejich reprezentaci ve formě formátů a dokumentů.

K dříve uvedenému seznamu systémů (obrázek 1.2-4) nyní přidáváme konkrétní typy formátů a dokumentů, se kterými často pracují:

■ Investor (CAPEX)

- Finanční údaje: rozpočty, prognózy výdajů (strukturované údaje).
- Údaje o tržních trendech: analýzy trhu (strukturovaná a nestrukturovaná data).
- Právní a smluvní údaje: smlouvy (textové údaje).



Obrázek 3.2-2 Ve stavebnictví se používá mnoho systémů s různými rozhraními, které pracují s různými typy dat.

Systémy řízení (PMS, CAFM, CQMS).

- Data projektu: grafy, úkoly (strukturovaná data).
- Údaje o údržbě zařízení: plány údržby (textové a polostrukturované údaje).
- Údaje o kontrole kvality: normy, inspekční zprávy (textová a nestrukturovaná data).

CAD, FEM a BIM

- Technické výkresy: architektonické, stavební plány (geometrická data, nestrukturovaná data).
- Stavební modely: 3D -modely, materiálová data (geometrická a polostrukturovaná data).
- Technické výpočty: analýza zatížení (strukturovaná data).

Systémy řízení staveniště (EHS, SCM)

- Bezpečnostní a zdravotní údaje: bezpečnostní protokoly (textové a strukturované údaje).
- Data dodavatelského řetězce: zásoby, objednávky (strukturovaná data).
- Denní výkazy: pracovní doba, produktivita (strukturovaná data).

Drony, AR/VR, GIS, 3D -printing

- Geodata: topografické mapy (geometrická a strukturovaná data).
- Data v reálném čase: video a fotografie (nestrukturovaná data).
- Modely pro 3D - tisk: digitální výkresy (geometrická data).

Další systémy řízení (4D BPM, 5D ERP1).

- Údaje o čase a nákladech: harmonogramy, odhady (strukturovaná data).
- Řízení změn: záznamy o změnách projektu (textová a strukturovaná data).
- Vykazování výkonnosti: ukazatele úspěšnosti (strukturované údaje).

Integrace dat a komunikace (CDE, RFID, AMS, RPM).

- Výměna dat: výměna dokumentů, datové modely (strukturovaná a textová data).
- RFID a sledování dat: logistika, správa majetku (strukturovaná data).
- Monitorování a kontrola: senzory na stavbách (strukturovaná a nestrukturovaná data).

Každý systém ve stavebnictví - od systémů pro řízení stavby až po provozní databáze - tedy pracuje s vlastním typem informací: strukturovanými, textovými, geometrickými a dalšími. "Datová krajina", se kterou musí odborníci denně pracovat, je nesmírně rozmanitá. Prostý výčet formátů však neodhalí složitost skutečné práce s informacemi.

V praxi se společnosti potýkají s tím, že data, i když jsou získána ze systémů, nejsou připravena k použití "tak, jak jsou". To platí zejména pro texty, obrázky, soubory PDF, CAD a další formáty, které je obtížné analyzovat pomocí standardních nástrojů. Proto je dalším klíčovým krokem transformace dat - proces, bez něhož nelze efektivně automatizovat zpracování, analýzu, vizualizaci a rozhodování.

Transformace dat: kritický základ moderní podnikové analýzy

Většina společností se dnes potýká s paradoxem: přibližně 80% jejich každodenních procesů se stále opírá o klasická strukturovaná data - známé tabulky Excelu a relační databáze (RDBMS) [66]. Zároveň je však 80% nových informací, které vstupují do digitálního ekosystému firem, nestrukturovaných nebo volně strukturovaných (obr. 3.2-3) [52]. Patří sem text, grafika, geometrie, obrázky, CAD -modely,

dokumentace ve formátu PDF, zvukové a obrazové záznamy, elektronická korespondence a mnoho dalšího.

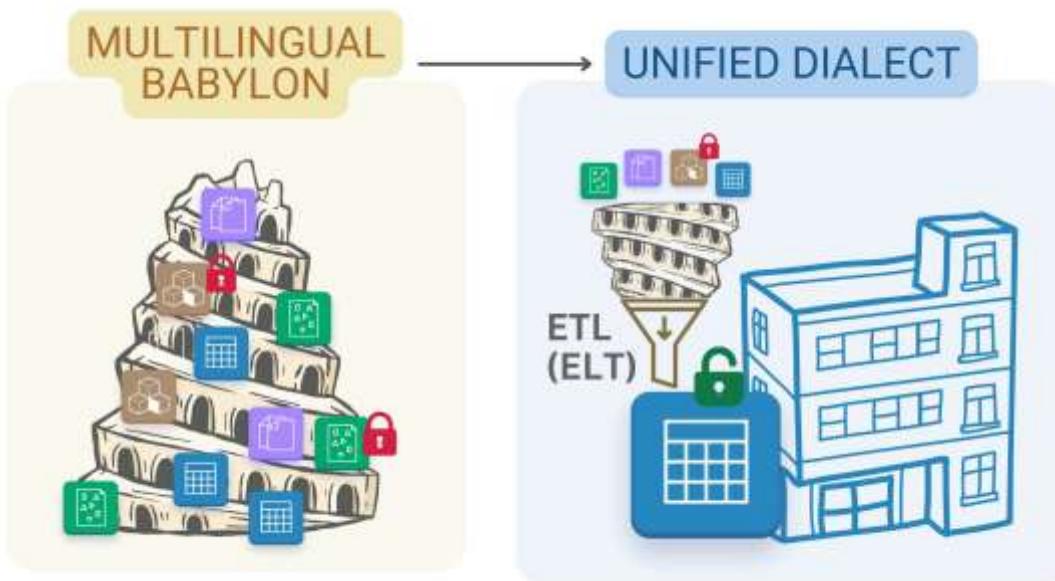
Objem nestrukturovaných dat navíc stále rychle roste - roční tempo růstu se odhaduje na 55-65% [67]. Taková dynamika způsobuje vážné potíže při integraci nových informací do stávajících podnikových procesů. Ignorování tohoto toku multiformátových dat vede ke vzniku informačních mezer a snižuje řiditelnost celého digitálního prostředí podniku.



Obr. 3.2-3 Každoroční nárůst nestrukturovaných dat vytváří problémy při integraci proudových informací do podnikových procesů.

Ignorování složitých nestrukturovaných a nepřehledných volně strukturovaných dat v procesech automatizace může vést ke značným mezerám v informačním prostředí společnosti. V dnešním světě nekontrolovatelného a lavinovitého pohybu informací musí společnosti přjmout hybridní přístup ke správě dat, který zahrnuje účinné metody pro práci se všemi typy dat.

Klíč k efektivní správě dat spočívá v organizaci, strukturování a klasifikaci různých typů dat "Babel" (včetně nestrukturovaných, textových a geometrických formátů, do strukturovaných nebo volně strukturovaných dat). Tento proces transformuje chaotické soubory dat do organizovaných struktur pro integraci do systémů, a tím umožňuje rozhodování na jejich základě (obrázek 3.2-4).



Obr. 3.2-4 Hlavním úkolem oddělení správy dat je převést "babylon" různorodých a různoformátových dat do strukturovaného a kategorizovaného systému.

Jednou z hlavních překážek takové harmonizace zůstává nízká úroveň interoperability mezi různými digitálními platformami - "sila", o nichž jsme hovořili v předchozích kapitolách.

Národní institut pro standardy a technologie (NIST, USA) podle zprávy zdůrazňuje [68], že špatná kompatibilita dat mezi různými stavebními platformami vede ke ztrátě informací a značným dodatečným nákladům. Jen v roce 2002 způsobily problémy s interoperabilitou softwaru v investiční výstavbě v USA ztráty v celkové výši 15,8 miliardy dolarů ročně, přičemž dvě třetiny těchto ztrát nesou vlastníci a provozovatelé budov, zejména při provozu a údržbě [68]. Studie také uvádí, že standardizace datových formátů může tyto ztráty snížit a zvýšit efektivitu v celém životním cyklu objektu.

Podle studie CrowdFlower z roku 2016 [69], která zahrnovala 16 000 datových vědců z celého světa, zůstávají hlavním problémem "špinavá" a multiformátová data. Podle této studie nejsou nejcennějším zdrojem výsledné databáze nebo modely strojového učení, ale čas strávený přípravou informací.

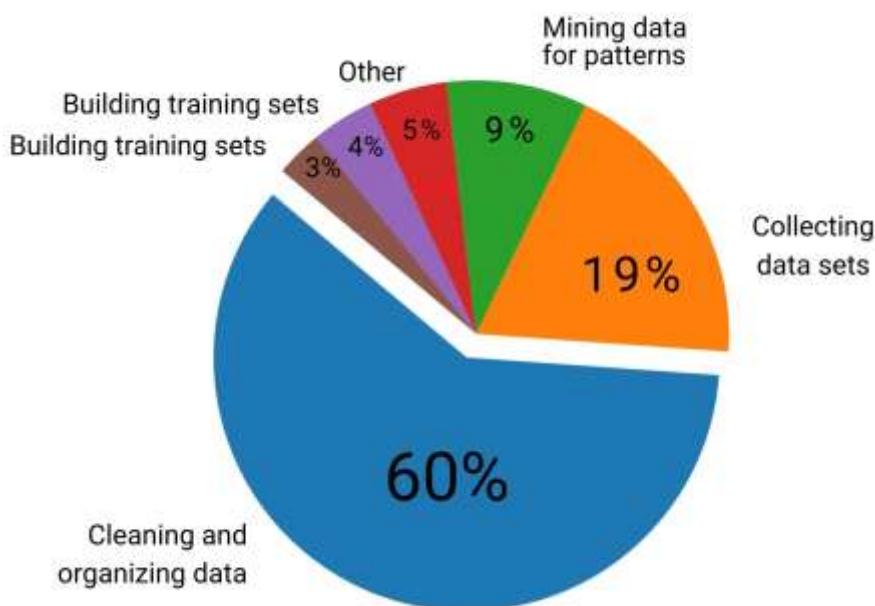
Čištění, formátování a organizování zabere až 60 procent času analytika a správce dat. Téměř pětinu zabere hledání a shromažďování správných datových souborů, které jsou často ukryti v silech a pro analýzu nedostupné. A jen asi 9 procent času je věnováno přímo modelování, analýze, vytváření předpovědí a testování hypotéz. Zbytek je věnován komunikaci, vizualizaci, reportování a vyhledávání podpůrných informačních zdrojů.

Průměrně je práce manažera s daty rozdělena takto (obrázek 3.2-5):

- **Čištění a uspořádání dat (60%):** Čistá a strukturovaná data mohou výrazně zkrátit pracovní dobu analytika a urychlit proces plnění úkolů.
- **Sběr dat (19%):** Hlavní výzvou pro odborníky na datovou vědu je vyhledávání relevantních datových souborů. Firemní data jsou často uložena v chaoticky uspořádaných "silech", což

ztěžuje přístup k potřebným informacím.

- **Modelování/strojové učení (9%):** Často brání nejasnost obchodních cílů ze strany zákazníků. Absence jasného zadání může zmařit potenciál i toho nejlepšího modelu.
- **Ostatní úkoly (5%):** kromě zpracování dat se analytici zabývají výzkumem, zkoumáním dat z různých úhlů pohledu, sdělováním výsledků prostřednictvím vizualizací a zpráv a doporučováním optimalizace procesů a strategií.



Obr. 3.2-5 Čemu věnují manažeři pracující s daty nejvíce času (podle [70]).

Tyto odhady potvrzují i další studie. Podle studie Xplenty publikované v časopise BizReport v roce 2015 [71] stráví 50 až 90% času profesionálů v oblasti business intelligence (BI) (BI) přípravou dat pro analýzu.

Čištění, validace a uspořádání dat představuje kritický základ pro všechny navazující datové a analytické procesy a zabere až 90% času datových vědců.

Tato pečlivá práce, která je pro koncového uživatele neviditelná, má zásadní význam. Chybě v nezpracovaných údajích nevyhnutelně zkreslují analýzy, jsou zavádějící a mohou vést k nákladným chybám v řízení. Proto se procesy čištění a standardizace dat - od odstranění duplicit a doplnění vynechaných údajů až po harmonizaci měrných jednotek a sladění se společným modelem - stávají základním kamenem moderní digitální strategie.

Důkladná transformace, čištění a standardizace dat tak nejen zabírá většinu času specialistů (až 80% práce s daty), ale také určuje možnost jejich efektivního využití v rámci moderních podnikových procesů. Samotnou organizací a čištěním dat se však úkol optimálního řízení informačních toků v podniku nevyčerpává. Ve fázi organizace a strukturování se stává volba vhodného datového modelu, který přímo ovlivňuje pohodlí a efektivitu práce s informacemi v dalších fázích zpracování.

Vzhledem k tomu, že data a obchodní cíle se liší, je důležité porozumět vlastnostem datových modelů

a umět vybrat nebo vytvořit správnou strukturu. V závislosti na stupni strukturovanosti a způsobu popisu vztahů mezi prvky existují tři hlavní modely: strukturovaný, volně strukturovaný a grafický. Každý z nich je vhodný pro různé úlohy a má své silné a slabé stránky.

Datové modely: vztahy v datech a vztahy mezi prvky

Data v informačních systémech jsou organizována různými způsoby - v závislosti na úkolech a požadavcích na ukládání, zpracování a přenos informací. Klíčovým rozdílem mezi typy datových modelů, tedy formou, v níž jsou informace uloženy, je míra strukturovanosti a způsob, jakým jsou popsány vztahy mezi prvky.

Strukturovaná data mají jasné a opakovatelné schéma: jsou uspořádána do tabulek s pevně danými sloupci. Tento formát zajišťuje předvídatelnost, snadné zpracování a efektivitu při provádění dotazů SQL, filtrování a agregaci. Příklady - databáze (RDBMS), Excel, CSV.

Volně strukturovaná data umožňují flexibilní strukturu: různé prvky mohou obsahovat různé sady atributů a mohou být uloženy jako hierarchie. Příkladem jsou formáty JSON, XML nebo jiné formáty dokumentů. Tato data jsou výhodná, pokud je třeba modelovat vnořené objekty a vztahy mezi nimi, ale na druhou stranu komplikují analýzu a standardizaci dat (obr. 3.2-6).

	Data Model	Storage Format	Example
	Relational	CSV, SQL	A table of doors in Excel
	Hierarchical	JSON, XML	Nested door objects inside a room
	Graph-based	RDF, GraphDB	Relationships between building elements

Obrázek 3.2-6 Datový model je logická struktura, která popisuje, jak jsou data v systému uspořádána, uložena a zpracovávána.

Volba vhodného formátu závisí na cílech:

- Pokud je důležitá rychlosť filtrování a analýzy - postačí relační tabulky (SQL, CSV, RDBMS, sloupcové databáze).
- Pokud je vyžadována flexibilita struktury - je lepší použít JSON nebo XML.
- Pokud mají data složité vztahy - grafové databáze poskytují přehlednost a škálovatelnost.

V klasických relačních databázích (RDBMS) je každá entita (např. dveře) reprezentována řádkem a její vlastnosti sloupci tabulky. Například tabulka položek z kategorie "Dveře" může obsahovat pole ID,

Výška, Šířka, Požární odolnost a ID místnosti označující místo (obr. 3.2-7).

V klasických relačních databázích (RDBMS) se vztahy vytvářejí ve formě tabulek, kde každý záznam představuje objekt a sloupce představují jeho parametry. V tabulkovém formátu vypadají údaje o dveřích v projektu takto, kde každý řádek představuje samostatný prvek - dveře s jejich jedinečným identifikátorem a atributy a spojení s místností je provedeno prostřednictvím parametru "ID místnosti".



Door ID	Room ID	Height (mm)	Width (mm)	Fireproof
ID1001	101	2000	900	Yes
ID1002	101	2100	800	No
ID1003	102	2000	850	Yes

Obr. 3.2-7 Informace o třech prvcích kategorie "Dveře" projektu v tabulkově strukturované podobě.

Ve volně strukturovaných formátech, jako je JSON nebo XML, jsou data uložena v hierarchické nebo vnořené podobě, kdy prvky mohou obsahovat další objekty a jejich struktura se může lišit. To umožňuje modelovat složité vztahy mezi prvky. Podobné informace o dveřích v projektu, které byly zaznamenány ve strukturované podobě (obrázek 3.2-7), jsou reprezentovány ve volně strukturovaném formátu (JSON) takovým způsobem (obrázek 3.2-8), že se stávají vnořenými objekty v rámci místností (Rooms - ID), což logicky odráží hierarchii.

1 {
2 "Rooms": [
3 {
4 "ID": 101,
5 "Doors": [
6 {"ID": 1, "Height": 2000, "Width": 900, "Fireproof": "Yes"},
7 {"ID": 2, "Height": 2100, "Width": 800, "Fireproof": "No"}
8]
9 },
10 {
11 "ID": 102,
12 "Doors": [
13 {"ID": 3, "Height": 2000, "Width": 850, "Fireproof": "Yes"}
14]
15 }
16]
17 }
18 }

Obr. 3.2-8 Informace o prvcích kategorie "Dveře" projektu ve formátu JSON.

V modelu grafu jsou data reprezentována jako uzly (vrcholy) a vazby (hrany) mezi nimi. To umožňuje vizualizovat složité vztahy mezi objekty a jejich atributy. V případě dat dveří a místností v projektu je reprezentace grafu následující:

- **Uzly (nodes)** představují hlavní entity: místnosti (Room 101, Room 102) a dveře (ID1001, ID1002, ID1003).
- **Žebra (odkazy)** ukazují vztahy mezi těmito entitami, např. příslušnost dveří k určité místnosti.
- **Atributy** jsou mapovány na uzly a obsahují vlastnosti entit (výška, šířka, požární odolnost dveří).



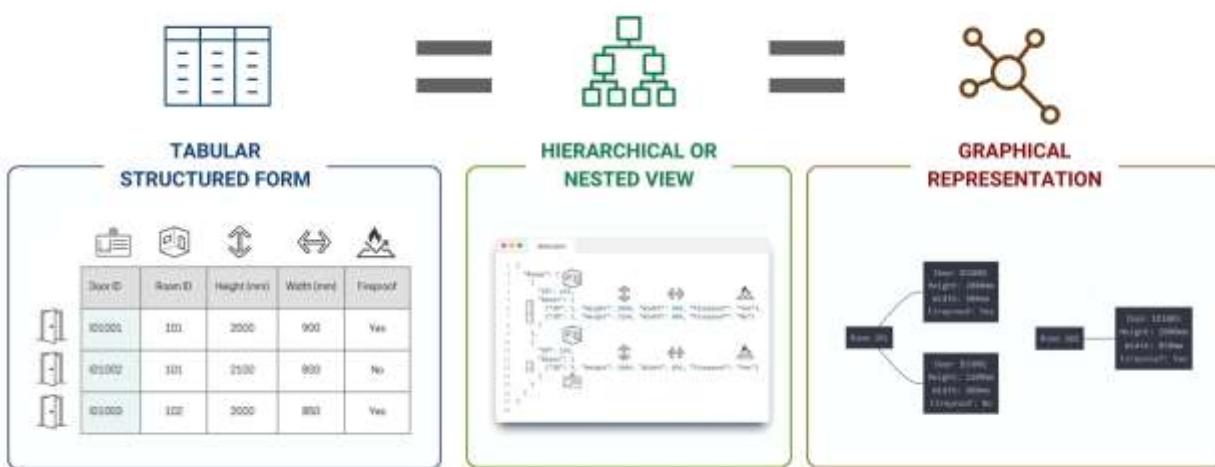
Obr. 3.2-9 Informace o entitě dveří projektu v grafickém zobrazení.

V grafovém datovém modelu popisu dveří jsou každá místnost a každé dveře samostatnými uzly.

Dveře jsou s místnostmi spojeny prostřednictvím hran, které označují, že dveře patří k určité místnosti. Atributy dveří (výška, šířka, požární odolnost) jsou uloženy jako vlastnosti příslušných uzelů. Podrobněji o grafových formátech a o tom, jak vznikla grafová sémantika ve stavebnictví, pojednává kapitola "Vznik sémantiky a ontologie ve stavebnictví".

Grafové databáze jsou efektivní v případech, kdy nejsou důležitá ani tak samotná data, ale vztahy mezi nimi, například v doporučovacích systémech, směrovacích systémech nebo při modelování složitých vztahů v projektech facility managementu. Grafový formát zjednoduší vytváření nových vztahů tím, že umožňuje přidávat do grafu nové datové typy, aniž by se měnila struktura úložiště. Ve srovnání s relačními tabulkami a strukturovanými formáty však v grafu neexistuje žádná další datová konektivita - přenesením dvourozměrných databázových dat do grafu se nezvýší počet vztahů a neumožnuje se získat nové informace.

Forma a schéma dat by měly být přizpůsobeny konkrétnímu případu použití a prováděným úkolům. Pro efektivní práci v podnikových procesech je důležité používat takové nástroje a takové datové modely, které pomáhají co nejrychleji a nejsnadněji získat výsledky.



Obrázek 3.2-10 Stejné informace o prvcích projektu mohou být uloženy v různých formátech pomocí různých datových modelů.

Většina velkých společností se dnes potýká s problémem nadměrné složitosti dat. Každá ze stovek nebo tisíců aplikací používá vlastní datový model, což vytváří nadměrnou složitost - jednotlivý model je často desítkykrát složitější, než je nutné, a souhrn všech modelů je tisíckrát složitější. Tato nadměrná složitost výrazně ztěžuje práci vývojářům i koncovým uživatelům.

Taková složitost představuje závažná omezení pro vývoj a údržbu systémů společnosti. Každý nový prvek v modelu vyžaduje dodatečný kód, implementaci nové logiky, důkladné testování a přizpůsobení stávajícím řešením. To vše zvyšuje náklady a zpomaluje práci automatizačního týmu ve firmě a mění i jednoduché úlohy v nákladné a časově náročné procesy.

Složitost ovlivňuje všechny úrovně datové architektury. V relačních databázích se projevuje rostoucím počtem tabulek a sloupců, často nadbytečných. V objektově orientovaných systémech se složitost zvyšuje množstvím tříd a vzájemně souvisejících vlastností. Ve formátech, jako je XML nebo JSON, se složitost projevuje nepřehlednými vnořenými strukturami, jedinečnými klíči a nekonzistentními schématy.

Přílišná složitost datových modelů způsobuje, že systémy jsou nejen méně efektivní, ale také obtížně srozumitelné pro koncové uživatele a v budoucnu i pro velké jazykové modely a agenty LLM. Právě problém s pochopením a složitostí datových modelů a zpracování dat vyvolává otázku: jak učinit data natolik jednoduchá, aby začala být skutečně rychle užitečná.

I když jsou datové modely zvoleny rozumně, jejich užitečnost se dramaticky snižuje, pokud je přístup k datům omezen. Proprietární formáty a uzavřené platformy brání integraci, komplikují automatizaci a odebírají kontrolu nad vlastními informacemi, čímž vytvářejí nejen silo nových dat, ale i uzamčené silo, ke kterému lze přistupovat pouze na základě autorizace dodavatele. Pro pochopení rozsahu problému je důležité zvážit, jak přesně uzavřené systémy ovlivňují digitální procesy ve stavebnictví.

Proprietární formáty a jejich dopad na digitální procesy

Jednou z klíčových výzev, kterým stavební společnosti při digitalizaci čelí, je omezený přístup k datům. To ztěžuje integraci systémů, snižuje kvalitu informací a znesnadňuje organizaci efektivních procesů. Přičinou těchto potíží je často používání proprietárních formátů a uzavřených softwarových řešení.

Bohužel dosud mnoho programů používaných ve stavebnictví umožňuje uživateli ukládat data pouze v proprietárních formátech nebo v cloudových úložištích, ke kterým lze přistupovat pouze prostřednictvím přísně omezených rozhraní. A nezřídka jsou tato řešení postavena v závislosti na ještě uzavřenějších systémech větších dodavatelů. V důsledku toho jsou i ti vývojáři, kteří by rádi nabízeli otevřenější architektury, nuceni dodržovat pravidla diktovaná velkými dodavateli.

Zatímco moderní systémy pro správu stavebních dat stále více podporují otevřené formáty a standardy (obr. 3.1-5), databáze založené na CAD (BIM) a související systémy ERP a CAFM zůstávají izolovanými proprietárními "ostrovky" v digitálním prostředí oboru (obr. 3.2-11).

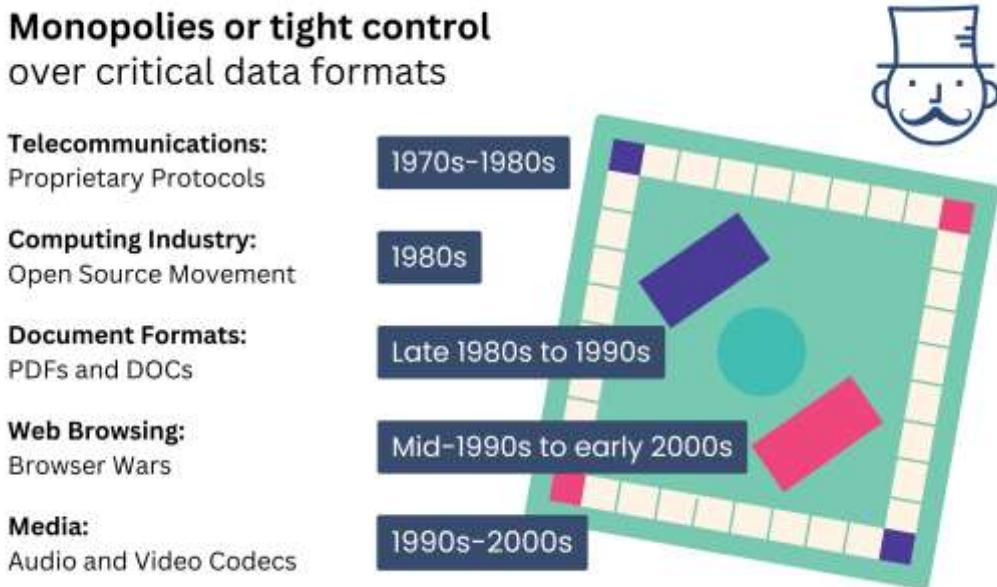


Obrázek 3.2-11 Uzavřená a proprietární povaha dat vytváří překážky pro jejich integraci a přístup k nim.

Uzavřené a monopolizované formáty a protokoly nejsou problémem pouze ve stavebnictví. V mnoha odvětvích hospodářství začal boj proti uzavřeným standardům a omezenému přístupu k datům zpomalením inovací (obrázek 3.2-12), existencí umělých překážek vstupu nových subjektů na trh a prohlubující se závislostí na velkých dodavatelích. S rychlým růstem významu dat orgány pro hospodářskou soutěž jednoduše nestačí reagovat na výzvy, které představují nové digitální trhy, a v důsledku toho se uzavřené formáty a omezený přístup k datům v podstatě stávají digitálními "hranicemi", které omezují tok informací a růst [63].

Pokud budou stroje vyrábět vše, co potřebujeme, bude naše situace záviset na tom, jak bude toto zboží distribuováno. Všichni si budou moci užívat života v blahobytu pouze tehdy, pokud bude bohatství vyrobené stroji sdíleno. Nebo většina lidí skončí v naprosté chudobě, pokud se majitelům automobilů podaří úspěšně lobbovat proti přerozdělování bohatství. Zatím se zdá, že se věci vyvíjejí druhým směrem, kdy technologie vedou ke stále větší nerovnosti [72].

- Stephen Hawking, astrofyzik, 2015



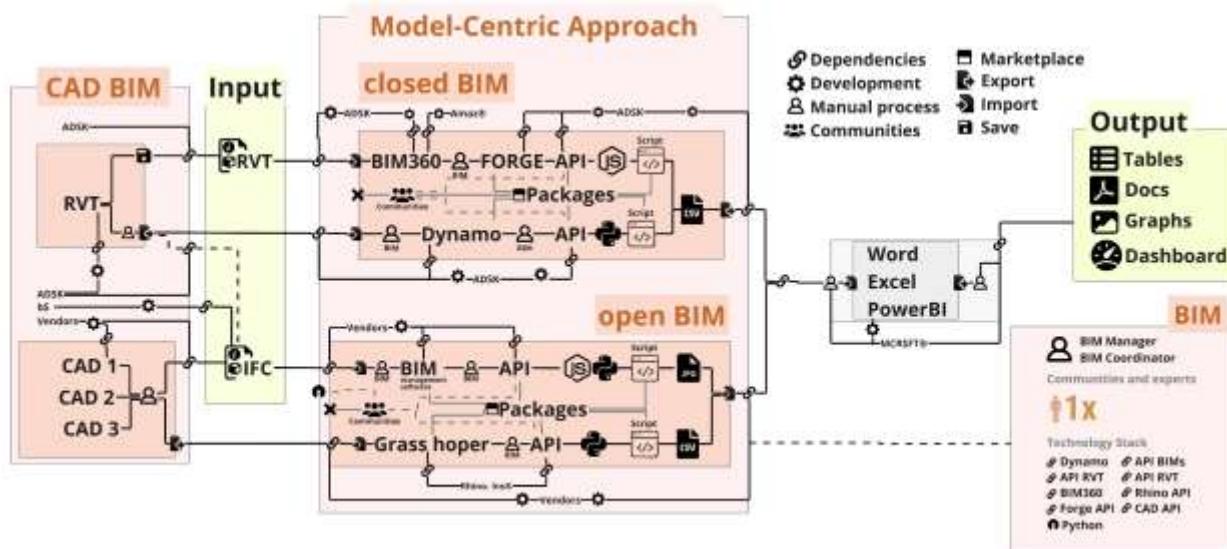
Obr. 3.2-12 Monopolní vlastnictví klíčových datových formátů a protokolů není výhradním problémem stavebnictví.

V důsledku uzavřeného přístupu k databázím programů se dnes datoví manažeři, datoví analytici, IT specialisté a vývojáři vytvářející aplikace pro přístup k datům, jejich zpracování a automatizaci ve stavebnictví potýkají s četnými závislostmi na dodavatelích softwaru (obr. 3.2-13). Tyto závislosti v podobě dalších přístupových vrstev vyžadují vytváření řešení se specializovanými API -připojeními a speciálními nástroji a softwarem.

Rozhraní API (Application Programming Interface) je formalizované rozhraní, jehož prostřednictvím může jeden program komunikovat s druhým a vyměňovat si data a funkce bez nutnosti přístupu ke zdrojovému kódu. Rozhraní API popisuje, jaké požadavky může externí systém zadávat, v jakém formátu by měly být a jaké odpovědi obdrží. Jedná se o standardizovanou "smlouvu" mezi softwarovými moduly.

Velké množství závislostí na uzavřených řešeních způsobuje, že se celá architektura kódu a logika obchodních procesů ve firmě stává "špagetovou architekturou" nástrojů závislých na politice dodavatele softwaru v oblasti poskytování kvalitního přístupu k datům.

Závislost na uzavřených řešeních a platformách vede nejen ke ztrátě flexibility, ale také ke skutečným obchodním rizikům. Změna licenčních podmínek, uzavření přístupu k datům, změna formátů nebo struktury API - to vše může zablokovat kritické procesy. Nejednou se ukáže, že aktualizace jedné tabulky vyžaduje přepracování celého bloku integrací a konektorů (obr. 3.2-13) a jakákoli rozsáhlá aktualizace softwaru nebo jeho dodavatele API se stává potenciální hrozbou pro stabilitu celého podnikového systému.



Obr. 3.2-13 Příklad velkého množství závislostí při zpracování CAD -dat vytváří překážky pro integraci dat v ekosystému stavebních firem.

Vývojáři a systémoví architekti jsou v takových podmínkách nuceni pracovat nikoliv pro očekávání, ale pro přežití. Místo toho, aby zaváděli nová řešení, přizpůsobují se. Místo vývoje se snaží udržet kompatibilitu. Místo automatizace a zrychlování procesů tráví čas studiem dalších uzavřených rozhraní, dokumentací API a nekonečným přestavováním kódu.

Práce s uzavřenými formáty a systémy není jen technickou výzvou, ale i strategickým omezením. Navzdory zjevným příležitostem, které nabízí moderní automatizace, umělá inteligence, LLM a prediktivní analýza, mnoho společností nedokáže plně využít jejich potenciál. A překážky, které staví proprietární formáty (obrázek 3.2-13), znemožňují podnikům přístup k vlastním datům. To je možná ironie digitální transformace ve stavebnictví.

Transparentnost dat a otevřené systémy nejsou luxusem, ale nezbytnou podmínkou rychlosti a efektivity. Bez otevřenosti jsou podnikové procesy plné zbytečné byrokracie, mnohovrstevnatých schvalovacích řetězců a rostoucí závislosti na principu HiPPO - rozhodování na základě názoru nejlépe placeného člověka.

Přesto se na obzoru rýsuje změna paradigmatu. Navzdory dominanci proprietárních řešení si stále více společností uvědomuje omezení architektur inspirovaných čtvrtou průmyslovou revolucí. Dnes se vektor posouvá směrem k principům páté revoluce, kde jsou v centru pozornosti data jako strategické aktivum, otevřená rozhraní (API) a skutečná interoperabilita mezi systémy.

Tento přechod znamená posun od uzavřených ekosystémů k flexibilním, modulárním digitálním architekturám, kde jsou klíčové otevřené formáty, standardy a transparentní výměna dat.

Otevřené formáty mění přístup k digitalizaci

Stavebnictví bylo jedním z posledních odvětví, které řešilo problém uzavřených a vlastnických dat. Na rozdíl od jiných hospodářských odvětví se zde digitalizace rozvíjela pomalu. Mezi důvody patří tradiční konzervativní povaha odvětví, převaha různorodých lokálních řešení a hluboce zakořeněná povaha správy dokumentů v papírové podobě. Po desetiletí se klíčové stavební procesy opíraly o fyzické výkresy, telefonáty a nesynchronizované databáze. V tomto kontextu byly uzavřené formáty dlouho vnímány spíše jako norma než jako překážka.

Zkušenosti z jiných průmyslových odvětví ukazují, že odstranění překážek v přístupu k uzavřeným datům vede k prudkému nárůstu inovací, zrychlení vývoje a zvýšení konkurence [73]. Ve vědě umožňuje výměna otevřených dat urychlit objevy a rozvíjet mezinárodní spolupráci. V medicíně může zlepšit účinnost diagnostiky a léčby. V softwarovém inženýrství - k vytvoření ekosystémů spolutvorby a rychlého zlepšování produktů.

Podle zprávy společnosti McKinsey "Open Data: Odemkněte inovace a produktivitu pomocí toku informací" 2013. [74] mají otevřená data potenciál uvolnit 3 až 5 bilionů dolarů ročně v sedmi klíčových odvětvích, včetně stavebnictví, dopravy, zdravotnictví a energetiky. Podle téže studie umožňují decentralizované datové ekosystémy velkým stavebním společnostem a dodavatelům snížit náklady na vývoj a údržbu softwaru, což urychluje zavádění digitálních technologií.

Přechod na otevřené architektury, který již dávno začal v jiných odvětvích ekonomiky, se postupně rozšiřuje i do stavebnictví. Velké společnosti a veřejní zadavatelé, a zejména finanční organizace, které kontrolují investice do stavebních projektů, stále více požadují využívání otevřených dat a přístup ke zdrojovému kódu výpočtů, kalkulací a aplikací. Od vývojářů se již neočekává pouze vytváření digitálních řešení a zobrazování konečných čísel projektu - očekává se od nich transparentnost, reprodukovatelnost a nezávislost na dodavatelích aplikací třetích stran.

Použití řešení s otevřeným zdrojovým kódem poskytuje zákazníkovi jistotu, že i když externí vývojáři přestanou spolupracovat nebo projekt opustí, nebude to mít vliv na možnost dalšího vývoje nástrojů a systémů. Jednou z hlavních výhod otevřených dat je jejich schopnost eliminovat závislost vývojářů aplikací na konkrétních platformách pro přístup k datům.

Pokud se společnost nemůže zcela vzdát proprietárních řešení, je možným kompromisem použití technik reverzního inženýrství. Tyto legální a technicky správné metody umožňují převést uzavřené formáty na přístupnější, strukturovanější a vhodné pro integraci. To je důležité zejména v případě potřeby napojení na starší systémy nebo migrace informací z jednoho softwarového prostředí do druhého.

Jedním z nejzářivějších příkladů v historii přechodu na otevřené formáty a uplatnění reverzního inženýrství (legálního nabourávání proprietárních systémů) ve stavebnictví je historie boje za otevření formátu DWG, hojně využívaného v systémech počítacem podporovaného projektování (CAD). V roce 1998 v reakci na monopol jednoho dodavatele softwaru vytvořilo dalších 15 dodavatelů CAD novou alianci nazvanou "Open DWG", aby poskytli vývojářům bezplatné a nezávislé nástroje pro práci s formátem DWG (de facto standard pro přenos výkresů) bez nutnosti používat proprietární software nebo uzavřené API. Tato událost byla zlomovým bodem, který umožnil desetitisícům firem získat bezplatný přístup k uzavřenému formátu oblíbeného CAD řešení od konce 80. let až do současnosti a vytvořit kompatibilní řešení, což podpořilo konkurenci na trhu CAD [75]. Dnes se SDK "Open DWG", který byl poprvé vytvořen již v roce 1996, používá téměř ve všech řešeních, v nichž je možné importovat, upravovat a exportovat formát DWG, mimo oficiální aplikace tvůrce formátu DWG.

Podobnou transformaci si vynucují i další technologičtí giganti. Společnost Microsoft, kdysi symbol proprietárního přístupu, otevřela zdrojový kód .NET Framework, začala používat Linux v infrastruktuře cloudové služby Azure a koupila GitHub, aby posílila svou pozici v komunitě open source. [76]. Společnost Meta (dříve Facebook) uvolnila modely umělé inteligence s otevřeným zdrojovým kódem, například řadu Llama, aby podpořila inovace a spolupráci při vývoji agentů umělé inteligence. Generální ředitel Mark Zuckerberg předpokládá, že platformy s otevřeným zdrojovým kódem budou v příštím desetiletí udávat směr technologického pokroku [77].

Open Source je model vývoje a distribuce softwaru, v němž je zdrojový kód otevřen k volnému použití, studiu, úpravám a distribuci.

Otevřená data a řešení s otevřeným zdrojovým kódem se stávají nejen trendem, ale i základem digitální udržitelnosti. Poskytují firmám flexibilitu, odolnost, kontrolu nad vlastními rozhodnutími a možnost škálovat digitální procesy bez závislosti na zásadách dodavatelů. A co je stejně důležité, vracejí podnikům kontrolu nad nejcennějším zdrojem 21. století - jejich daty.

Změna paradigmatu: Open Source jako konec éry dominance dodavatelů softwaru

Stavebnictví prochází změnou, kterou nelze vyčíslit obvyklým způsobem. Koncept přístupu založeného na datech, orientovaného na data a využívání nástrojů Open Source vede k přehodnocení pravidel hry, na kterých stojí softwaroví giganti na trhu.

Na rozdíl od předchozích technologických transformací nebude tento přechod aktivně podporován dodavateli. Změna paradigmatu ohrožuje jejich tradiční obchodní modely založené na licencích, předplatném a konzultacích. Nová realita nezahrnuje hotový produkt nebo placené předplatné - vyžaduje změnu nastavení procesů a myšlení.

Pro správu a vývoj řešení datových center založených na otevřených technologiích budou muset společnosti přehodnotit interní procesy. Specialisté z různých oddělení budou muset nejen spolupracovat, ale také přehodnotit způsob své spolupráce.

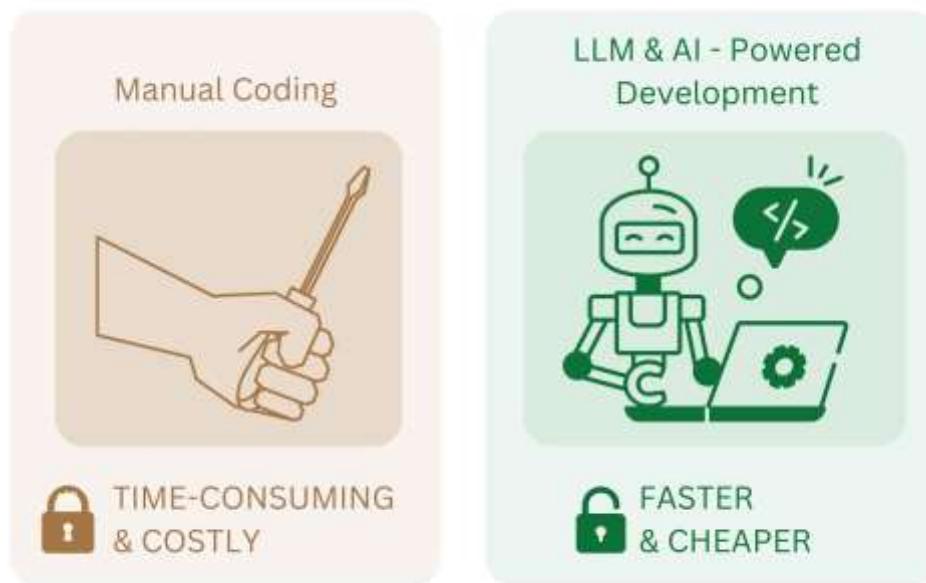
Nové paradigma předpokládá využívání otevřených dat a řešení s otevřeným zdrojovým kódem, kde zvláštní roli při vytváření softwarového kódu budou hrát nástroje založené na umělé inteligenci a velkých jazykových modelech (LLM), nikoli programátoři. Již v polovině roku 2024 bude více než 25% nového kódu ve společnosti Google vytvořeno pomocí umělé inteligence [78]. V budoucnu bude kódování pomocí LLM provádět 80% práce za pouhých 20% času (obrázek 3.2-14).

Podle studie McKinsey 2020 [79] nahrazují GPU v analytice stále více CPU díky svému vysokému výkonu a podpoře moderních nástrojů s otevřeným zdrojovým kódem. To umožňuje firmám zrychlit zpracování dat bez významných investic do drahého softwaru nebo najímání nedostatkových specialistů.

Přední poradenské firmy, jako jsou McKinsey, PwC a Deloitte, zdůrazňují rostoucí význam otevřených standardů a aplikací s otevřeným zdrojovým kódem v různých odvětvích.

Podle zprávy PwC Open Source Monitor 2019 [80] 69% společností se 100 a více zaměstnanci vědomě používá řešení Open Source. OSS je aktivně využíván zejména ve velkých firmách: Ve firmách s 200-499 zaměstnanci je to 71%, v kategorii 500-1999 zaměstnanců 78% a mezi firmami s více než 2000 zaměstnanci až 86%. Podle zprávy Synopsys OSSRA 2023 obsahovalo 96% analyzovaných kódových bází komponenty s otevřeným zdrojovým kódem [81].

Úkolem vývojáře v budoucnu nebude ručně psát kód, ale navrhovat datové modely, architektury toků a řídit agenty umělé inteligence, kteří na požádání vytvoří správné výpočty. Uživatelská rozhraní budou minimalistická a interakce bude založena na dialogu. Klasické programování ustoupí návrhu na vysoké úrovni a orchestraci digitálních řešení (obrázek 3.2-14). Současné trendy - jako jsou platformy s nízkým kódem (obr. 7.4-6) a ekosystémy s podporou LLM (obr. 7.4-4) - výrazně sníží náklady na vývoj a údržbu IT systémů.



Obr. 3.2-14 Zatímco dnes jsou aplikace vytvářeny ručně programátory, v budoucnu bude významná část kódu generována pomocí umělé inteligence a řešení založených na LLM.

Tento přechod bude odlišný od předchozích a velcí dodavatelé softwaru pravděpodobně nebudou jeho katalyzátorem.

Studie Harvard Business School "The Value of Open Source Software" 2024 [40] odhaduje celkovou hodnotu open source softwaru ze dvou hledisek. Na jedné straně, pokud bychom spočítali, kolik by stálo vybudování všech existujících open source řešení od nuly, jednalo by se o částku přibližně 4,15 miliardy dolarů. Na druhou stranu, pokud si představíme, že každá firma vyvíjí vlastní obdobu Open Source řešení sama (což se děje všude), aniž by měla přístup k existujícím nástrojům, pak by celkové náklady na podnikání dosáhly kolosálních 8,8 bilionu dolarů - to je cena poptávky.

Není těžké uhodnout, že žádný významný dodavatel softwaru nemá zájem na zmenšení softwarového trhu s potenciální hodnotou 8,8 bilionu dolarů na pouhých 4,15 miliardy dolarů. To by znamenalo snížit objem poptávky více než 2000krát. Taková transformace je pro dodavatele, jejichž obchodní modely jsou postaveny na dlouholetém udržování závislosti zákazníků na uzavřených řešeních, jednoduše nerentabilní. Společnosti, které očekávají, že jim někdo nabídne pohodlné a otevřené řešení na klíč, tak mohou být zklamány - tito dodavatelé se prostě neobjeví.

Přechod na otevřenou digitální architekturu neznamená ztrátu pracovních míst ani příjmů. Naopak, vytváří podmínky pro flexibilní a přizpůsobivé obchodní modely, které mohou nakonec vytlačit tradiční trh s licencemi a krabicovým softwarem.

Místo prodeje licencí - služby, místo uzavřených formátů - otevřené platformy, místo závislosti na dodavateli - nezávislost a možnost vytvářet řešení pro skutečné potřeby. Ti, kteří dříve nástroje pouze

používali, se budou moci stát jejich spoluautory. A ti, kteří umí pracovat s daty, modely, scénáři a logikou, se ocitnou v centru nové digitální ekonomiky průmyslu. O těchto změnách a o tom, jaké nové role, obchodní modely a formáty spolupráce se kolem otevřených dat objevují, budeme více hovořit v závěrečné, desáté části knihy.

Řešení založená na otevřených datech a otevřeném kódu umožní firmám soustředit se na efektivitu obchodních procesů, nikoli na boj se zastaralými rozhraními API a integraci uzavřených systémů. Vědomý přechod na otevřenou architekturu může výrazně zvýšit produktivitu a snížit závislost na dodavatelích.

Přechod do nové reality neznamená jen změnu přístupů k vývoji softwaru, ale také přehodnocení samotného principu práce s daty. V centru této proměny není kód, ale informace: její struktura, dostupnost a interpretovatelnost. Zde se do popředí dostávají otevřená a strukturovaná data, která se stávají nedílnou součástí nové digitální architektury.

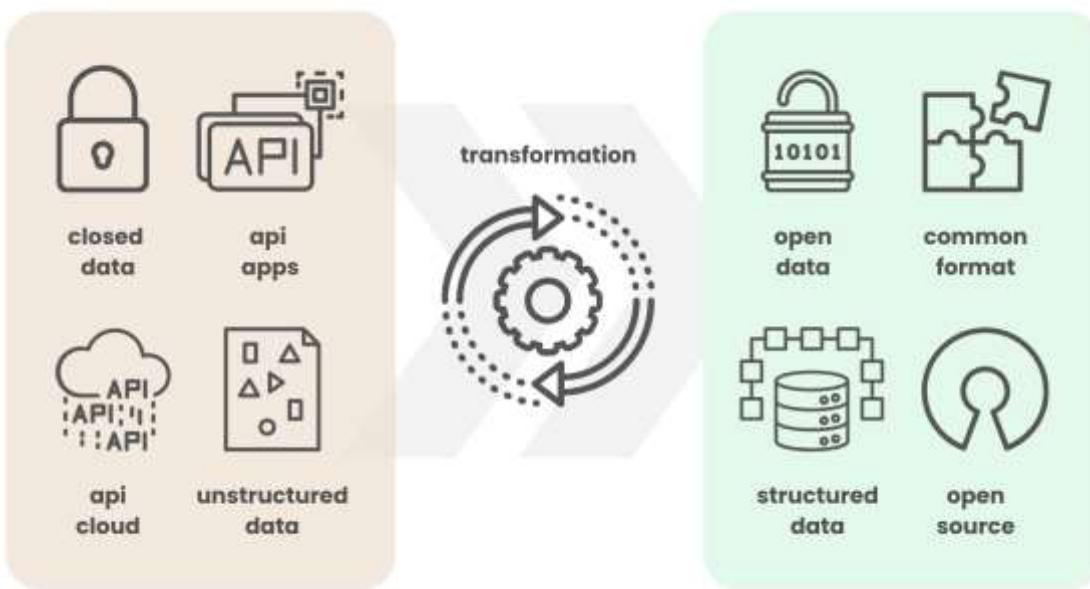
Strukturovaná otevřená data: základ digitální transformace

Zatímco v minulých desetiletích rozhodoval o udržitelnosti podnikání především výběr softwarových řešení a závislost na konkrétních dodavatelích, v dnešní digitální ekonomice je klíčovým faktorem kvalita dat a schopnost s nimi efektivně pracovat. Otevřený kód je důležitou součástí nového technologického paradigmatu, ale jeho potenciál se skutečně uvolní pouze tehdy, když jsou data srozumitelná, uspořádaná a strojově čitelná. Mezi všemi typy datových modelů se základním kamenem udržitelné digitální transformace stávají strukturovaná otevřená data.

Hlavní výhodou strukturovaných otevřených dat je jednoznačná interpretace a možnost automatizovaného zpracování. To umožňuje výrazné zvýšení efektivity jak na úrovni jednotlivých operací, tak v rámci celé organizace.

Podle zprávy společnosti Deloitte "Proces přenosu dat při transformaci podniku". [82] je spolupráce s IT při řízení přenosu strukturovaných dat klíčová. Podle zprávy britské vlády "Data Analytics and AI in Government Project Delivery" (2024) [83] je odstranění překážek pro sdílení dat mezi různými projekty a organizacemi klíčem ke zvýšení efektivity řízení projektů. Dokument zdůrazňuje, že standardizace datových formátů a zavedení zásad otevřených dat může zabránit duplicitě informací, minimalizovat časové ztráty a zlepšit přesnost prognóz.

Pro stavebnictví, kde tradičně převládá vysoká míra rozdílnosti a různorodosti formátů, hráje strukturovaný sjednocovací proces a strukturovaná otevřená data zásadní roli při utváření ucelených a zvládnutelných procesů (obr. 4.1-14). Umožňují účastníkům projektu soustředit se spíše na zvyšování produktivity než na řešení technických problémů spojených s nekompatibilitou uzavřených platform, datových modelů a formátů.



Obr. 3.2-15 Otevřená strukturovaná data snižují závislost na softwarových řešeních a platformách a urychlují inovace.

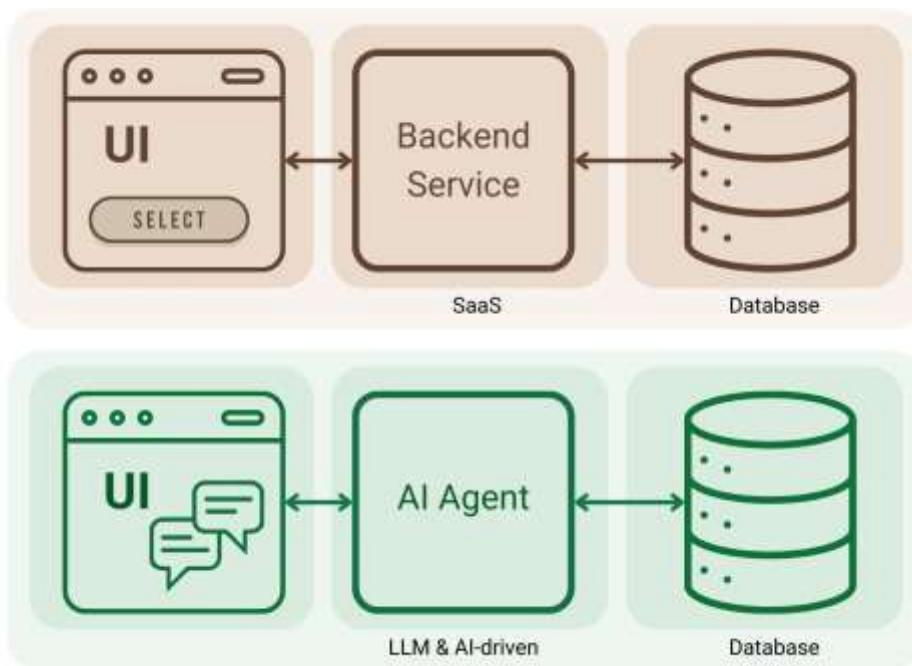
Moderní technologické nástroje, o nichž se budeme podrobněji zmínovat dále v knize, umožňují nejen shromažďovat informace, ale také je automaticky čistit: odstraňovat duplicity, opravovat chyby a normalizovat hodnoty. To znamená, že analytici a inženýři nepracují s různorodými dokumenty, ale s uspořádanou znalostní bází vhodnou pro analýzu, automatizaci a rozhodování.

Udělejte to co nejjednodušší, ale ne jednodušší.

- Albert Einstein, teoretický fyzik (autorství citátu je sporné [84])

Většinu uživatelských rozhraní pro práci s daty lze dnes vytvářet automaticky - bez nutnosti ručního psaní kódu pro každý obchodní případ. To vyžaduje infrastrukturní vrstvu, která rozumí struktuře dat, modelu a logice bez dalších instrukcí (obrázek 4.1-15). Tento přístup umožňuje právě strukturovaná data: formuláře, tabulky, filtry a pohledy lze generovat automaticky s minimálním programátorským úsilím.

Nejdůležitější rozhraní, která jsou pro uživatele kritická, mohou stále vyžadovat ruční zdokonalení. Ve většině případů - a to je 50 až 90 procent pracovních scénářů - však postačí automatické generování aplikací a výpočtů bez použití speciálních aplikací pro tento účel (obr. 3.2-16), což výrazně snižuje náklady na vývoj a údržbu, omezuje chybovost a urychluje implementaci digitálních řešení.



Obr. 3.2-16 Architektonické modely pro práci s daty: tradiční architektura aplikace a model orientovaný na umělou inteligenci s LLM.

Dalším krokem v digitální evoluci je přechod od architektur postavených na jednotlivých aplikacích k intelligentně řízeným systémům založeným na jazykových modelech (LLM). V takové architektuře se strukturovaná data stávají nejen předmětem ukládání, ale také základem pro interakci s nástroji umělé inteligence, které jsou schopny analyzovat, interpretovat a doporučovat akce na základě kontextu.

V následujících kapitolách se podíváme na reálné příklady implementace architektury založené na otevřených strukturovaných datech a ukážeme si, jak se jazykové modely používají k automatické interpretaci, ověřování a zpracování dat. Tyto praktické případy vám pomohou lépe pochopit, jak nová digitální logika funguje v praxi - a jaké výhody přináší společnostem, které jsou připraveny na transformaci.



KAPITOLA 3.3.

LLM A JEJICH ÚLOHA PŘI ZPRACOVÁNÍ DAT A OBCHODNÍCH PROCESECH

Chatovací místnosti LLM: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok k automatizaci datových procesů.

Vznik velkých jazykových modelů (Large Language Models, LLM) byl přirozeným pokračováním hnutí za strukturovaná otevřená data a filozofie otevřeného softwaru. Když se data stanou uspořádanými, přístupnými a strojově čitelnými, dalším krokem je nástroj, který dokáže s těmito informacemi pracovat, aniž by bylo nutné psát složitý kód nebo mít specializované technické znalosti.

LLM jsou přímým produktem otevřenosti: rozsáhlé otevřené datové sady, publikace a hnutí Open Source. Bez otevřených vědeckých článků, veřejně dostupných textových dat a kultury společného vývoje by neexistoval ChatGPT ani jiné LLM. LLM je v jistém smyslu "destilátem" nahromaděných digitálních znalostí lidstva, shromážděných a vzdělaných díky principům otevřenosti.

Moderní velké jazykové modely (LLM - Large Language Models), jako jsou ChatGPT ® (OpenAI), LlaMa™ (Meta AI), Mistral DeepSeek™, Grok™ (xAI), Claude™ (Anthropic), QWEN™, poskytují uživatelům možnost formulovat dotazy na data v přirozeném jazyce. To zpřístupňuje práci s informacemi nejen vývojářům, ale i analytikům, inženýrům, plánovačům, manažerům a dalším odborníkům, kterým bylo programování dříve vzdálené

LLM (Large Language Model) je umělá inteligence, která je vyškolena k porozumění a generování textu na základě obrovského množství dat shromážděných z celého internetu. Dokáže analyzovat kontext, odpovídat na otázky, vést dialog, psát texty a generovat softwarový kód.

Jestliže dříve vizualizace, zpracování nebo analýza dat vyžadovaly znalost speciálního programovacího jazyka: Python, SQL, R nebo Scala, stejně jako schopnost pracovat s knihovnami jako Pandas, Polars nebo DuckDB a mnoha dalšími, od roku 2023 se situace radikálně mění. Nyní může uživatel jednoduše popsat, co chce získat - a model sám vygeneruje kód, provede jej, zobrazí tabulku nebo graf a vysvětlí výsledek. Poprvé po desetiletích se vývoj technologií nevydal cestou komplikací, ale radikálního zjednodušení a zpřístupnění.

Tento princip - "zpracovávat data pomocí slov (výzev)". - znamenal novou etapu ve vývoji práce s informacemi a fakticky posunul tvorbu řešení na ještě vyšší úroveň abstrakce. Stejně jako kdysi již nebylo nutné, aby uživatelé rozuměli technickým základům internetu, aby mohli provozovat internetové obchody nebo vytvářet webové stránky pomocí WordPress, Joomla a dalších modulárních systémů s otevřeným zdrojovým kódem (av tor knihy pracuje s těmito systémy od roku 2005, včetně vzdělávacích

a inženýrských online platform). - To následně vedlo k rozmachu digitálního obsahu a online podnikání - inženýři, analytici a manažeři dnes mohou automatizovat pracovní postupy bez znalosti programovacích jazyků. Tomu napomáhají výkonné LLM - bezplatné i open source, jako jsou LLaMA, Mistral, Qwen, DeepSeek a další - které zpřístupňují pokročilé technologie nejširšímu publiku.

Velké jazykové modely LLM: jak to funguje

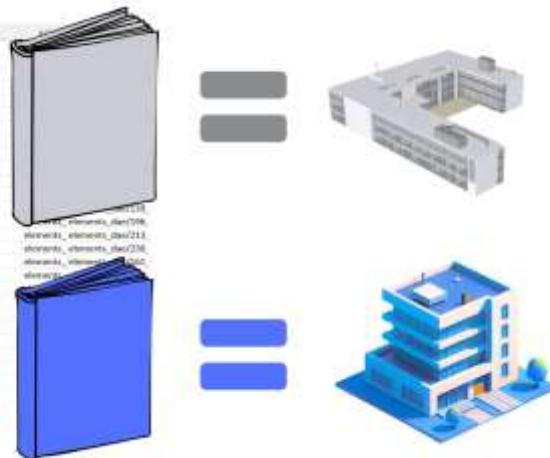
Velké jazykové modely (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok) jsou neuronové sítě vyškolené na obrovském množství textových dat z internetu, knih, článků a dalších zdrojů. Jejich hlavním úkolem je porozumět kontextu lidské řeči a generovat smysluplné odpovědi.

Moderní LLM je založen na architektuře Transformer, kterou navrhli výzkumníci společnosti Google v roce 2017 [85]. Klíčovou součástí této architektury je mechanismus pozornosti, který modelu umožňuje zohledňovat vztahy mezi slovy bez ohledu na jejich pozici v textu.

Proces učení LLM je vzdáleně podobný způsobu, jakým se lidé učí jazyk, jen je milionkrát větší. Model analyzuje miliardy příkladů slov a výrazů a identifikuje vzory ve struktuře jazyka a v logice sémantických přechodů. Celý text je rozčleněn na tokeny - minimální sémantické jednotky (slova nebo jejich části), které jsou následně transformovány do vektorů ve vícerozměrném prostoru (obr. 8.2-2). Tyto vektorové reprezentace umožňují stroji "pochopit" skryté vztahy mezi pojmy, místo aby jednoduše pracoval s textem jako s posloupností symbolů.

Velké jazykové modely nejsou jen nástrojem pro generování textu. Dokážou rozpoznat význam, najít souvislosti mezi pojmy a pracovat s daty, i když jsou prezentována v různých formátech. Hlavní je, aby informace byly rozděleny do srozumitelných modelů a reprezentovány jako tokeny, se kterými může LLM pracovat.

Stejný přístup lze použít i u stavebních projektů. Pokud si projekt představíme jako druh textu, kde každá budova, prvek nebo konstrukce je symbolem, můžeme začít zpracovávat tyto informace podobným způsobem. Stavební projekty lze přirovnat ke knihám, které jsou uspořádány do kategorií, kapitol a skupin odstavců sestávajících z minimálních tokenů - prvků stavebního projektu (obrázek 3.3-1). Převodem datových modelů do strukturovaného formátu můžeme strukturovaná data převést také do vektorových bází (obr. 8.2-2), které jsou ideálním zdrojem pro strojové učení a technologie, jako je LLM.



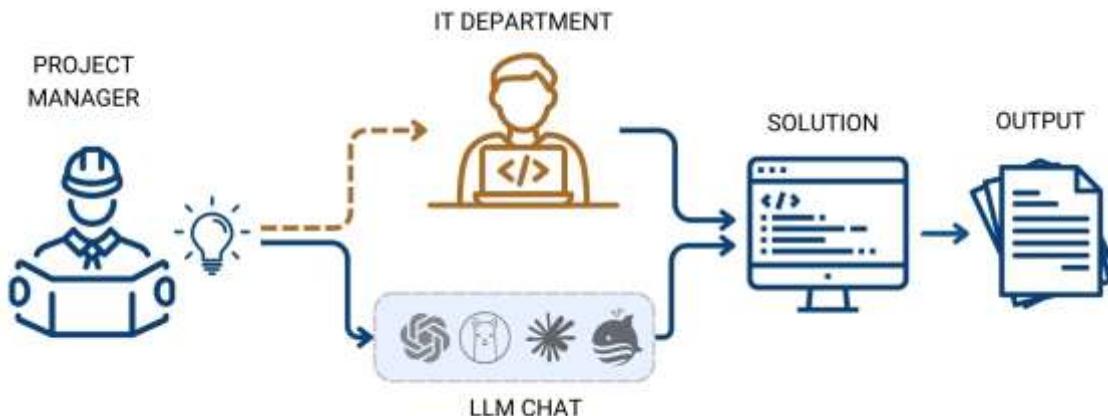
Obr. 3.3-1 Prvek stavebního projektu je jako token v textu: minimální jednotka, ze které se tvoří skupiny (odstavce) oddílů (kategorií) celého projektu.

Pokud je stavební projekt digitalizován a jeho prvky jsou reprezentovány jako tokeny nebo vektory, je možné k nim přistupovat spíše v přirozeném jazyce než prostřednictvím rigidních formálních dotazů. Zde se projevuje jedna z klíčových výhod LLM - schopnost pochopit význam dotazu a propojit jej s příslušnými daty.

Inženýr již nemusí psát SQL -dotaz nebo kód v jazyce Python, aby získal požadovaná data - může jednoduše, s pochopením LLM a struktury dat, formulovat úlohu obvyklým způsobem: "Najděte všechny železobetonové konstrukce s třídou betonu vyšší než B30 a vypočítejte jejich celkový objem". Model rozpozná význam dotazu, převede jej do strojově čitelné podoby, vyhledá data (seskupí a transformuje) a vrátí konečný výsledek.

Dokumenty, tabulky, modely projektů jsou převedeny do vektorové reprezentace (vložení) a uloženy do databáze. Když uživatel položí dotaz, dotaz je rovněž převeden na vektor a systém vyhledá nejrelevantnější data. To umožňuje systému LLM spoléhat se nejen na vlastní natrénované znalosti, ale také na skutečná podniková data, i když se objevila již po skončení trénování modelu.

Jednou z nejdůležitějších výhod LLM ve stavebnictví je možnost generovat programový kód. Namísto předávání technické úlohy programátorovi mohou odborníci úlohu popsat v přirozeném jazyce a model vytvoří potřebný kód, který lze použít (zkopírováním z chatu) při tvorbě kódu pro automatizaci procesu. LLM -modely umožňují specialistům bez hlubších znalostí programování přispět k automatizaci a zdokonalení podnikových procesů.



Obrázek 3.3-2 LLM poskytují uživatelům možnost psát kód a získávat výsledky bez nutnosti znalosti programování.

Podle studie provedené společností Wakefield Research a sponzorované společností SAP v roce 2024 [36], která se dotazovala 300 vedoucích pracovníků ve společnostech s ročními příjmy alespoň 1 miliardu USD v USA: 52% vedoucích pracovníků důvěřuje umělé inteligenci při analýze dat a poskytování doporučení pro rozhodování. Dalších 48% využívá AI k identifikaci dříve nezohledněných rizik a 47% využívá AI k navrhování alternativních plánů. Kromě toho 40% využívá AI pro vývoj nových produktů, plánování rozpočtu a průzkum trhu. Studie také ukázala pozitivní dopad AI na osobní život: 39% respondentů uvádí zlepšení rovnováhy mezi pracovním a soukromým životem, 38% zlepšení duševního zdraví a 31% nižší úroveň stresu.

Přes veškerou svou sílu však zůstávají LLM nástrojem, který je důležité používat vědomě. Jako každá technologie mají svá omezení. Jedním z nejznámějších problémů jsou takzvané "halucinace" - případy, kdy model s jistotou vyprodukuje věrohodnou, ale fakticky nesprávnou odpověď. Je proto velmi důležité pochopit, jak model funguje: jaká data a datové modely dokáže bezchybně interpretovat, jak interpretuje dotazy a odkud získává informace. Je také třeba mít na paměti, že znalosti LLM jsou omezeny datem jeho školení a bez napojení na externí data nemusí model zohledňovat aktuální normy, standardy, ceny nebo technologie.

Řešením těchto problémů je pravidelná aktualizace vektorových databází, připojení k relevantním zdrojům a vývoj autonomních agentů s umělou inteligencí, kteří nejen odpovídají na otázky, ale proaktivně využívají data pro školení, řídí úkoly, identifikují rizika, navrhují možnosti optimalizace a sledují výkonnost projektu.

Přechod na LLM -rozhraní ve stavebnictví není jen technologickou novinkou. Je to změna paradigmatu, která odstraňuje bariéry mezi lidmi a daty. Jde o možnost pracovat s informacemi stejně snadno, jako spolu mluvíme - a přitom získávat přesné, ověřené a použitelné výsledky.

Ty společnosti, které začnou tyto nástroje používat dříve než ostatní, získají významnou konkurenční výhodu. Ta zahrnuje urychlení práce, snížení nákladů a zlepšení kvality konstrukčních řešení díky

rychlému přístupu k analýze dat a schopnosti rychle najít odpovědi na složité otázky. Je však třeba vzít v úvahu také otázky bezpečnosti. Využívání cloudových služeb LLM -služeb může být spojeno s riziky úniku dat. Organizace proto stále častěji hledají alternativní řešení, která jim umožní nasadit nástroje LLM ve vlastní infrastruktuře - lokálně, s plnou ochranou a kontrolou nad informacemi.

Využití místních LLM pro citlivé firemní údaje

Vznik prvních chat-LLM v roce 2022 znamenal novou etapu ve vývoji umělé inteligence. Bezprostředně po širokém rozšíření těchto modelů však vyvstala oprávněná otázka: Jak bezpečné je přenášet data a dotazy související s firmou do cloutu? Většina cloudových jazykových modelů ukládala historii komunikace a nahrané dokumenty na své servery a pro společnosti, které pracují s citlivými informacemi, to představovalo vážnou překážku pro přijetí umělé inteligence.

Jedním z nejudržitelnějších a nejlogičtějších řešení tohoto problému bylo nasazení otevřeného zdrojového kódu LLM lokálně, v rámci podnikové IT infrastruktury. Na rozdíl od cloudových služeb fungují lokální modely bez připojení k internetu, nepřenášejí data na externí servery a poskytují firmám plnou kontrolu nad informacemi

Nejlepší otevřený model [Open Source LLM] je v současné době výkonnostně srovnatelný s uzavřenými modely [například ChatGPT, Claude], ale se zpožděním přibližně jednoho roku [77].

- Ben Cottier, vedoucí výzkumný pracovník neziskové výzkumné organizace Epoch AI, 2024

Velké technologické společnosti začaly zpřístupňovat své LLM pro místní použití. Příkladem přechodu k otevřené architektuře byla open source řada LLaMA společnosti Meta a rychle se rozvíjející čínský projekt DeepSeek. Vedle nich uvolnily výkonné modely osvobozené od omezení proprietárních platform také společnosti Mistral a Falcon. Tyto iniciativy nejen urychlily vývoj globální umělé inteligence, ale také poskytly společnostem, které dbají na ochranu soukromí, skutečné alternativy nezávislosti, flexibility a dodržování bezpečnostních předpisů.

Ve firemním prostředí, zejména ve stavebnictví, není ochrana dat jen otázkou pohodlí, ale i dodržování právních předpisů. Práce se zadávací dokumentací, odhadů, výkresy a důvěrnou korespondencí vyžaduje přísnou kontrolu. A právě zde poskytuje místní LLM potřebnou jistotu, že data zůstanou uvnitř firemního perimetru.

	Cloud LLMs (OpenAI, Claude)	Local LLMs (DeepSeek, LLaMA)
Data Control	Data is transmitted to third parties	Data remains within the company's network
License	Proprietary, paid	Open-source (Apache 2.0, MIT)
Infrastructure	Requires internet	Operates in an isolated environment
Customization	Limited	Full adaptation to company needs
Cost	Pay-per-token/request	One-time hardware investment + maintenance costs
Scalability	Easily scalable with cloud resources	Scaling requires additional local hardware
Security & Compliance	Risk of data leaks, may not meet strict regulations (GDPR, HIPAA)	Full compliance with internal security policies
Performance & Latency	Faster inference due to cloud infrastructure	Dependent on local hardware, may have higher latency
Integration	API-based integration, requires internet access	Can be tightly integrated with on-premise systems
Updates & Maintenance	Automatically updated by provider	Requires manual updates and model retraining
Energy Consumption	Energy cost is covered by provider	High power consumption for inference and training
Offline Availability	Not available without an internet connection	Works completely offline
Inference Cost	Pay-per-use model (cost scales with usage)	Fixed cost after initial investment

Obr. 3.3-3 Místní modely poskytují kompletní kontrolu a zabezpečení, zatímco clouдовá řešení nabízejí snadnou integraci a automatické aktualizace.

Hlavní výhody místního otevřeného softwaru LLM:

- Úplná kontrola nad daty. Veškeré informace zůstávají v rámci společnosti, čímž se eliminuje neoprávněný přístup a únik dat.
- Provoz offline. Žádná závislost na internetovém připojení, což je důležité zejména pro izolované IT infrastruktury. Tím je zajištěn nepřetržitý provoz i v případě sankcí nebo zablokování cloudových služeb.
- Flexibilita použití. Model lze použít pro generování textu, analýzu dat, psaní programového kódu, podporu návrhu a řízení obchodních procesů.
- Přizpůsobení podnikovým cílům. LLM lze školit na interních dokumentech, což umožňuje zohlednit specifika práce firmy a její oborové zvláštnosti. Místní LLM lze propojit s

platformami CRM, ERP nebo BI, což umožňuje automatizovat analýzu zákaznických dotazů, tvorbu reportů nebo dokonce předpovídání trendů.

Nasazení bezplatného modelu DeepSeek s otevřeným zdrojovým kódem -R1-7B na serveru pro přístup celého týmu uživatelů za 1000 USD měsíčně může potenciálně stát méně než roční poplatky za cloudová rozhraní API, jako je ChatGPT nebo Claude, a umožňuje společnostem získat plnou kontrolu nad svými daty, eliminuje jejich přenos na internet a pomáhá splnit regulační požadavky, jako je GDPR

V jiných odvětvích již místní LLM mění svůj přístup k automatizaci. V oblasti podpůrných služeb reagují na časté dotazy zákazníků, čímž snižují pracovní zátěž operátorů. V personálních odděleních analyzují životopisy a vybírají vhodné kandidáty. V elektronickém obchodě generují personalizované nabídky, aniž by odhalovaly údaje o uživatelích.

Podobný efekt se očekává i v odvětví stavebnictví. Integrací LLM s projektovými daty a standardy je možné urychlit přípravu dokumentace, automatizovat přípravu odhadů a prediktivních analýz nákladů. Zvláště slibnou oblastí se stává využití LLM ve spojení se strukturovanými tabulkami a datovými rámcemi.

Plná kontrola nad AI ve firmě a způsob nasazení vlastního LLM

Moderní nástroje umožňují firmám lokálně nasadit velký jazykový model (LLM) během několika hodin. To poskytuje úplnou kontrolu nad daty a infrastrukturou, eliminuje závislost na externích cloudových službách a minimalizuje riziko úniku informací. Toto řešení má význam zejména pro organizace pracující s citlivou projektovou dokumentací nebo obchodně citlivými údaji.

V závislosti na úlohách a zdrojích jsou k dispozici různé scénáře nasazení, od hotových řešení až po flexibilnější a škálovatelné architektury. Jedním z nejjednodušších nástrojů je Ollama, který umožňuje spouštět jazykové modely doslova na jedno kliknutí, aniž byste potřebovali hluboké technické znalosti. Rychlý začátek s aplikací Ollama:

1. Stáhněte si distribuci pro svůj operační systém (Windows / Linux / MacOS) z oficiálních stránek: ollama.com
2. Nainstalujte model pomocí příkazového řádku. Například pro model *Mistral*:

```
ollama run mistral
```

3. Po spuštění modelu je připraven k práci - můžete odesílat textové dotazy prostřednictvím terminálu nebo jej integrovat do jiných nástrojů. Spusťte model a proveděte dotaz:

ollama run mistral "Jak vytvořit výpočet se všemi zdroji pro práci na instalaci sádrokartonové příčky o šířce 100 mm?".

Pro ty, kteří dávají přednost práci ve známém vizuálním prostředí, je tu LM Studio, bezplatná aplikace s rozhraním připomínajícím ChatGPT

- Nainstalujte si LM Studio stažením distribuční sady z oficiálních webových stránek - lmstudio.ai.
- Prostřednictvím vestavěného katalogu vyberte model (např. Falcon nebo GPT-Neo-X) a stáhněte jej.
- Práce s modelem prostřednictvím intuitivního rozhraní připomínajícího ChatGPT, které je však zcela lokalizované.

	Developer	Parameters	GPU Requirements (GB)	Features	Best For
Mistral 7B	Mistral AI	7	8 (FP16)	Fast, supports multimodal tasks (text + images), fully open-source code	Lightweight tasks, mobile devices, laptops
LLaMA 2	Meta	7–70	16–48 (FP16)	High text generation accuracy, adaptable for technical tasks, CC-BY-SA license	Complex analytical and technical tasks
Baichuan 7B/13B	Baichuan Intelligence	7–13	8–16 (FP16)	Fast and efficient, great for large data processing, fully open-source code	Data processing, automating routine tasks
Falcon 7B/40B	Technology Innovation Institute (TII)	7–40	8–32 (FP16)	Open-source, high performance, optimized for fast work	Workloads with limited computational resources
DeepSeek-V3	DeepSeek	671	1543 (FP16) / 386 (4-bit)	Multilingual, 128K token context window, balanced speed and accuracy	Large enterprises, SaaS platforms, multitasking scenarios
DeepSeek-R1-7B	DeepSeek	7	18 (FP16) / 4.5 (4-bit)	Retains 92% of R1 capabilities in MATH-500, local deployment support	Budget solutions, IoT devices, edge computing

Obrázek 3.3-4 Srovnání populárních lokálních open source LLM -modelů.

Volba modelu závisí na požadavcích na rychlosť, přesnosť a dostupné hardwarové možnosti (obrázek 3.3-4). Malé modely, jako jsou Mistral 7B a Baichuan 7B, jsou vhodné pro nenáročné úlohy a mobilní zařízení, zatímco výkonné modely, jako je DeepSeek -V3, vyžadují značné výpočetní prostředky, ale nabízejí vysoký výkon a podporu více jazyků. V příštích letech bude trh s LLM rychle růst - budeme se setkávat se stále více lehkými a specializovanými modely. Místo univerzálních LLM pokryvajících

veškerý lidský obsah se budou objevovat modely vyškolené na základě úzkých doménových znalostí. Můžeme například očekávat vznik modelů určených výhradně pro zpracování inženýrských výpočtů, stavebních odhadů nebo dat ve formátech CAD. Takové specializované modely budou rychlejší, přesnější a bezpečnější - zejména v profesionálním prostředí, kde je důležitá vysoká spolehlivost a hloubka tématu.

Po spuštění místní aplikace LLM ji lze přizpůsobit konkrétním úkolům společnosti. K tomuto účelu se používá technika jemného doladění, kdy se model dále trénuje na interních dokumentech, technických pokynech, šablonách smluv nebo projektové dokumentaci.

RAG: Inteligentní LLM -asistenti s přístupem k firemním datům.

Další fází vývoje aplikace LLM v podnikání je integrace modelů se skutečnými podnikovými daty v reálném čase. Tento přístup se nazývá RAG (Retrieval-Augmented Generation) - vyhledávání s rozšířením. V této architektuře se jazykový model nestává pouhým dialogovým rozhraním, ale plnohodnotným inteligentním asistentem schopným navigovat v dokumentech, výkresech, databázích a poskytovat přesné, kontextové odpovědi.

Hlavní výhodou modelu RAG je možnost využívat interní údaje společnosti bez nutnosti předem trénovat model při zachování vysoké přesnosti a flexibilitě při zpracování informací.

Technologie RAG kombinuje dvě hlavní složky:

- **Vyhledávání:** model se připojuje k datovým úložištěm - dokumentům, tabulkám, souborům PDF - souborům, výkresům - a vyhledává příslušné informace podle požadavků uživatele.
- **Rozšířené generování:** na základě získaných dat model vygeneruje přesnou a informovanou odpověď, přičemž zohlední kontext a specifičnost dotazu.

Chcete-li spustit program LLM s podporou RAG, je třeba provést několik kroků:

- **Příprava dat:** shromáždění potřebných dokumentů, výkresů, specifikací, tabulek. Mohou být v různých formátech a strukturách, od PDF až po Excel.
- **Indexování a vektorizace:** pomocí nástrojů, jako je LlmalIndex nebo LangChain, se data převádějí do vektorové reprezentace, která umožňuje vyhledávat sémantické vazby mezi textovými fragmenty (více o vektorových databázích a převodu velkých polí do vektorové reprezentace, včetně projektů CAD, v části 8).
- **Dotazování asistenta:** po nahrání dat můžete modelu klást otázky a on bude hledat odpovědi v podnikovém rámci, nikoli v obecných znalostech získaných z internetu.

Předpokládejme, že společnost má složku constructionsite_docs, kde jsou uloženy smlouvy, pokyny, odhady a tabulky. Pomocí skriptu v jazyce Python (obr. 3.3-5) můžeme tuto složku prohledat a vytvořit vektorovou indexaci: každý dokument bude převeden na sadu vektorů odrážejících

sémantický obsah textu. Tím se z dokumentů stane jakási "mapa významů", na které se model může efektivně pohybovat a nacházet souvislosti mezi termíny a frázemi.

Model si například "pamatuje", že slova "vrácení" a "reklamace" se často vyskytují v části smlouvy týkající se přepravy materiálu na staveniště. Pokud je pak položena otázka - například "Jaká je naše lhůta pro vrácení?", může se stát, že se objeví otázka, která se týká vrácení materiálu. (Obr. 3.3-5 - řádek 11 kódu) - LLM analyzuje interní dokumenty a najde přesné informace, přičemž se chová jako inteligentní asistent schopný číst a porozumět obsahu všech firemních souborů.



```

1 from llama_index import SimpleDirectoryReader, VectorStoreIndex
2
3 # Load documents from the folder
4 documents = SimpleDirectoryReader("constructionsite_docs").load_data()
5
6 # Creating a vector index for semantic search
7 index = VectorStoreIndex.from_documents(documents)
8
9 # Integration with LLM (e.g. Llama 3)
10 query_engine = index.as_query_engine()
11 response = query_engine.query("What are the return terms in the contracts?")
12 print(response)

```

Obr. 3.3-5 LM čte složku se soubory - podobně jako když ji člověk otevře a vyhledá požadovaný dokument

Kód lze spustit na libovolném počítači s nainstalovaným jazykem Python. O používání jazyka Python a IDE pro spouštění kódu si povíme více v příští kapitole.

Místní nasazení LLM není jen trendem, ale strategickým řešením pro společnosti, které si cení bezpečnosti a flexibility. Nasazení LLM, ať už na lokálních firemních počítačích, nebo pomocí online řešení, je však pouze prvním krokem. Aby bylo možné aplikovat schopnosti LLM na reálné úlohy, musí společnosti používat nástroje, které jim umožní nejen přijímat odpovědi z chatu, ale také ukládat vytvořenou logiku ve formě kódu, který lze spustit mimo kontext používání LLM. To je důležité pro škálování řešení - správně uspořádané procesy umožňují aplikovat vývoj AI na několik projektů najednou nebo dokonce v rámci celé společnosti.

V této souvislosti hraje důležitou roli výběr vhodného vývojového prostředí (IDE). Moderní programovací nástroje umožňují nejen vyvíjet řešení založená na LLM, ale také je integrovat do stávajících podnikových procesů a proměnit je v automatizované ETL-Pipeline



KAPITOLA 3.4.

IDE S PODPOROU LLM A BUDOUCÍ PROGRAMOVÉ ZMĚNY

Výběr IDE: od experimentů LLM po podniková řešení

Při pronikání do světa automatizace, analýzy dat a umělé inteligence - zejména při práci s rozsáhlými jazykovými modely (LLM) - je velmi důležité zvolit správné integrované vývojové prostředí (IDE). Toto IDE bude vaším hlavním pracovním nástrojem: místem, kde se bude spouštět kód generovaný LLM, a to jak na místním počítači, tak v rámci podnikové sítě. Výběr IDE určuje nejen pohodlí vaší práce, ale také to, jak rychle budete schopni přejít od experimentálních dotazů LLM k plnohodnotným řešením začleněným do reálných podnikových procesů.

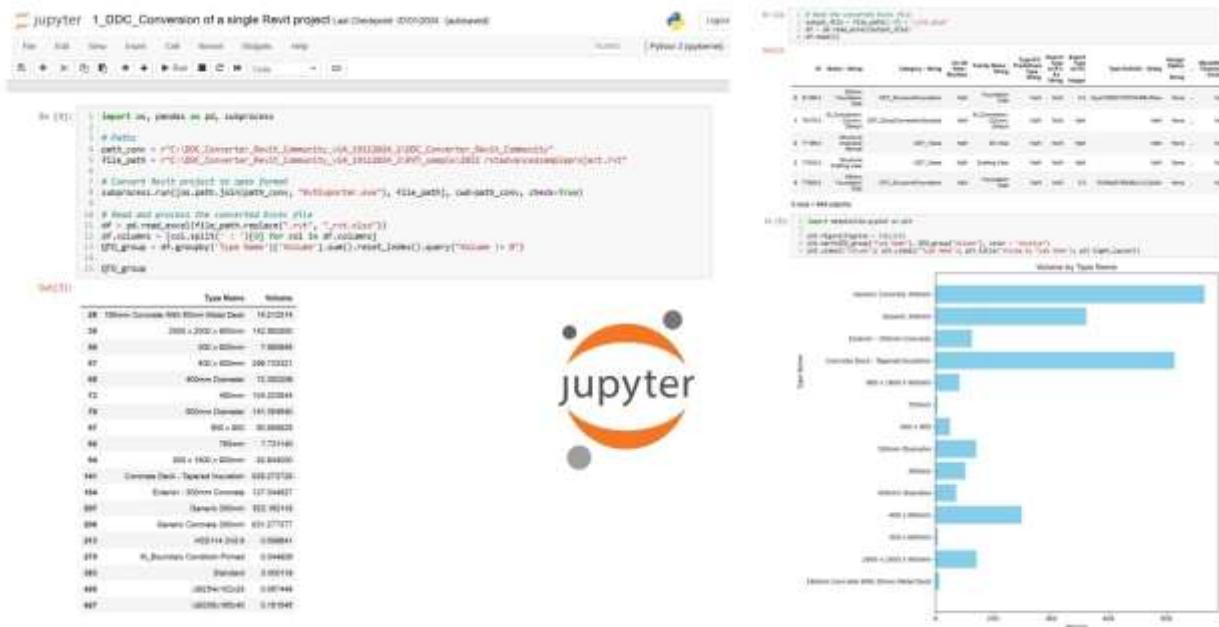
IDE (Integrated Development Environment) je univerzální stavební prvek počítače pro automatizaci procesů a zpracování dat. Místo toho, abyste měli pilu, kladivo, vrtačku a další nástroje odděleně, máte k dispozici jedno zařízení, které zvládne vše - rezat, upevňovat, vrtat a dokonce i kontrolovat kvalitu materiálů. IDE pro programátory je jednotný prostor, kde můžete psát kód (analogicky se stavbou - vytvářet výkresy), testovat jeho práci (sestavení stavebního modelu), hledat chyby (podobně jako kontrola pevnosti konstrukcí při stavbě) a spouštět hotový projekt (uvedení domu do provozu).

Přehled oblíbených IDE:

- **PyCharm®** (JetBrains) je výkonné profesionální IDE pro Python. Díky velkému množství vestavěných funkcí je vhodné pro seriózní projekty. Základní podpora interaktivních souborů Jupyter (IPYNB) je však k dispozici pouze v placené verzi a pro začátečníky může být rozhraní nepřehledné.

Soubor s příponou IPYNB (Interactive Python Notebook) je formát pro interaktivní zápisníky Jupyter® (obrázek 3.4-1), kde jsou kód, vizualizace a vysvětlivky spojeny do jednoho dokumentu. Tento formát je ideální pro vytváření sestav, analytických a výukových scénářů.

- **VS Code®** (Microsoft) je rychlý, flexibilní a přizpůsobitelný nástroj s bezplatnou podporou IPYNB a mnoha pluginy. Je vhodný pro začátečníky i profesionály. Umožňuje integraci zásuvných modulů GitHub Copilot a jazykového modelu, takže je skvělou volbou pro projekty AI a datové vědy.
- **Jupyter Notebook** - Klasická a oblíbená volba pro experimentování a učení. Umožňuje psát kód, přidávat vysvětlivky a vizualizovat výsledky v jediném rozhraní (obr. 3.4-1). Ideální pro rychlé testování hypotéz, práci s LLM a vytváření reprodukovatelných kroků analýzy divokých dat. Pro správu závislostí a knihoven doporučujeme používat Anaconda Navigator, vizuální rozhraní pro správu prostředí Python.



Obr. 3.4-1 Jupyter Notebook jeden z nejvhodnějších a nejoblíbenějších nástrojů pro vytváření procesů Pipeline.

■ **Google Collab™** (a platforma Kaggle (obrázek 9.2-5)) je cloudová alternativa k Jupyteru, která poskytuje bezplatný přístup ke GPU/TPU. Je to skvělé řešení pro začátek - žádná instalace lokálního softwaru a možnost pracovat přímo z prohlížeče. Podporuje integraci s Diskem Google a nově i s Gemini (LLM společnosti Google).

	PyCharm	VS Code	Jupyter Notebook	Google Colab
Complexity	High	Medium	Low	Low
.ipynb support	Paid	Free	Built-in	Built-in
Copilots	Yes	Yes	Yes	Yes
Computing resources	Local	Local	Local	Cloud
For whom	Professionals	Universal	Beginners	Experimenters

Obrázek 3.4-2 Srovnání IDE: Jupyter Notebook jeden z nejvhodnějších a nejjednodušších nástrojů pro vytváření procesů Pipeline.

Výběr IDE závisí na vašich úkolech. Pokud chcete rychle začít pracovat s umělou inteligencí, vyzkoušejte Jupyter Notebook nebo Google Collab. Pro seriózní projekty je lepší použít PyCharm nebo VS Code. Hlavní je začít pracovat. Moderní nástroje vám umožní rychle proměnit vaše experimenty ve

funkční řešení.

Všechna výše popsaná IDE umožňují vytvářet pipelines pro zpracování dat - tedy řetězce modulů bloků kódu (které mohou být generovány pomocí LLM), z nichž každý je zodpovědný například za jinou fázi:

- analytické scénáře,
- řetězce získávání informací z dokumentů,
- automatické reakce na základě RAG,
- generování zpráv a vizualizací.

Díky modulární struktuře lze každý krok představit jako samostatný blok: načítání dat → filtrování → analýza → vizualizace → export výsledků. Tyto bloky lze opakovaně používat, - upravovat a skládat do nových řetězců jako konstruktor, jen pro data.

Inženýrům, manažerům a analytikům se tak otevírá možnost dokumentovat logiku rozhodování ve formě kódu, který lze generovat pomocí LLM. Tento přístup pomáhá urychlit rutinní úkoly, automatizovat typické operace a vytvářet opakovatelné procesy, kde je každý krok jasné zdokumentován a transparentní pro všechny členy týmu.

Automatizované ETL Pipelines (obr. 7.2-3), Apache Airflow (obr. 7.4-4), Apache NiFi (obr. 7.4-5) a n8n (obr. 7.4-6) nástroje pro sestavování logických bloků pro automatizaci procesů budou podrobněji probrány v 7. a 8. části knihy.

IDE s podporou LLM a budoucí programové změny

Integrace umělé inteligence do vývojových procesů mění prostředí programování. Moderní prostředí už nejsou jen textové editory se zvýrazňováním syntaxe - mění se v inteligentní asistenty schopné pochopit logiku projektu, doplnit kód, a dokonce vysvětlit, jak konkrétní fragment kódu funguje. Na trhu se objevují produkty, které využívají umělou inteligenci k posunutí hranic běžného vývoje:

- **GitHub Copilot** (integrace s VS Code, PyCharm): AI -asistent, který generuje kód na základě komentářů nebo částečných popisů a mění textové návodů na hotová řešení.
- **Cursor** (fork VS Code s AI -kernel): umožňuje nejen dokončit psaní kódu, ale také klást otázky projektu, hledat závislosti a učit se z kódové základny.
- **JetBrains AI Assistant**: zásuvný modul pro JetBrains IDE (včetně PyCharm) s funkcí vysvětlování složitého kódu, optimalizace a vytváření testů.
- **Amazon CodeWhisperer**: obdoba Copilotu se zaměřením na bezpečnost a podporu služeb Amazon AWS.

Programování projde v nadcházejících letech dramatickou změnou. Hlavní důraz se přesune z rutinního psaní kódu na návrh modelů a datové architektury - vývojáři se budou více podílet na návrhu systému, zatímco AI převeze šablonovité úlohy: generování kódu, testy, dokumentaci a základní funkce. Budoucnost programování spočívá ve spolupráci lidí a AI, kdy stroje převezmou technickou rutinu a lidé se zaměří na kreativitu.

Programování v přirozeném jazyce se stane každodenní záležitostí. Personalizace IDE dosáhne nové

úrovňě - vývojová prostředí se naučí přizpůsobovat stylu práce uživatele a jeho firmám tím, že budou předvídat vzory, nabízet kontextová řešení a učit se z předchozích projektů

Tím se úloha vývojáře nezruší, ale radikálně se změní: od psaní kódu k řízení znalostí, kvality a procesů. Tento vývoj se dotkne i business intelligence, kde se tvorba reportů, vizualizací a aplikací pro podporu rozhodování bude stále častěji provádět prostřednictvím generování kódu a logiky s pomocí AI a LLM, chatu a agentových rozhraní.

Jakmile společnost nastaví chaty LLM a vybere vhodné vývojové prostředí, je dalším důležitým krokem uspořádání dat. Tento proces zahrnuje extrakci informací z různorodých zdrojů, jejich vyčištění, transformaci do strukturované podoby a integraci do podnikových systémů.

V moderním datově orientovaném přístupu ke správě dat je hlavním cílem přenést data do jediné univerzální podoby, která je kompatibilní s velkým množstvím nástrojů a aplikací. Ke zpracování procesů strukturování a strukturovaných dat jsou zapotřebí specializované knihovny. Jednou z nejvýkonnějších, nejflexibilnějších a nejoblíbenějších je knihovna Pandas pro Python. Umožňuje pohodlně zpracovávat tabulková data: filtrovat, seskupovat, čistit, připojovat, agregovat a sestavovat reporty.

Python Pandas: nepostradatelný nástroj pro práci s daty

Pandas zaujímá ve světě analýzy a automatizace dat zvláštní místo. Jedná se o jednu z nejoblíbenějších a nejpoužívanějších knihoven programovacího jazyka Python [86], která je určena pro práci se strukturovanými daty.

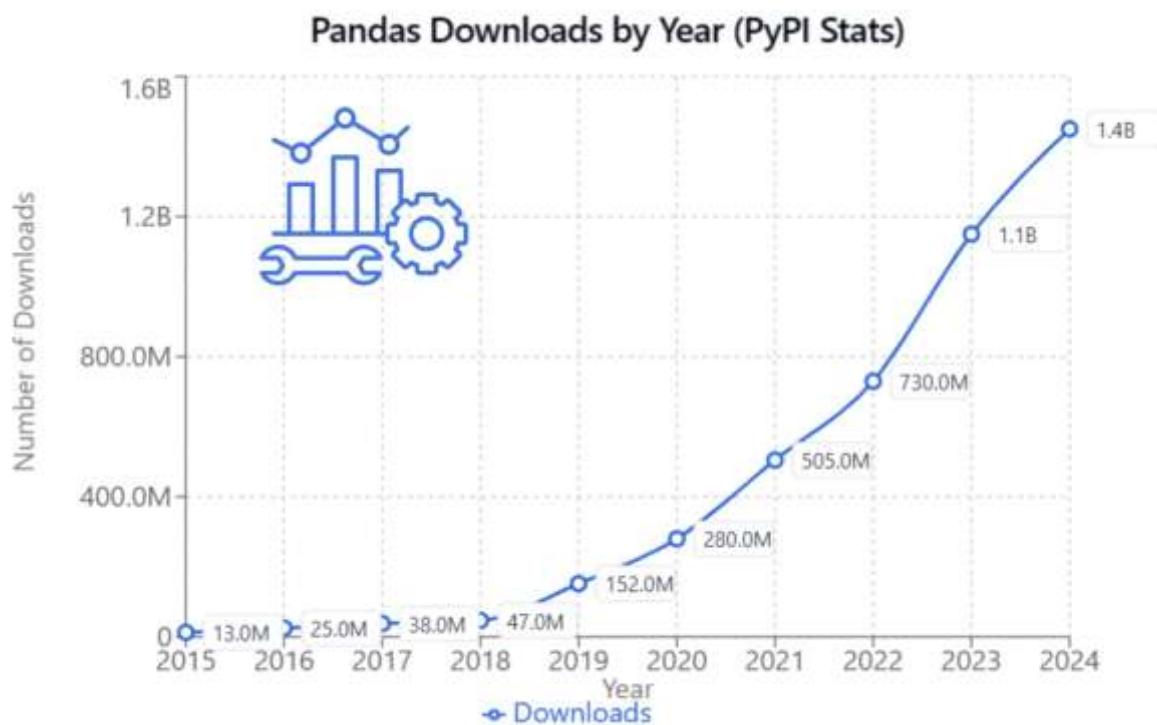
Knihovna je jako sada hotových nástrojů: funkcí, modulů, tříd. Stejně jako na stavbě nemusíte pokaždé vymýšlet kladivo nebo vodováhu, tak i v programování vám knihovny umožňují rychle řešit problémy, aniž byste museli znova vymýšlet základní funkce a řešení.

Pandas je open source knihovna jazyka Python , která poskytuje vysoce výkonné a intuitivní datové struktury, zejména DataFrame, univerzální formát pro práci s tabulkami. Pandas je švýcarský nůž pro analytiky, inženýry a vývojáře pracující s daty.

Python je vysokoúrovňový programovací jazyk s jednoduchou syntaxí, který se aktivně používá v analytice, automatizaci, strojovém učení a při vývoji webových stránek. Jeho popularita je dána čitelností kódu, multiplatformností a bohatým ekosystémem knihoven. K dnešnímu dni bylo pro Python vytvořeno více než 137 000 balíčků s otevřeným zdrojovým kódem [87] a toto číslo téměř denně roste. Každá taková knihovna je jakýmsi úložištěm hotových funkcí: od jednoduchých matematických operací až po složité nástroje pro zpracování obrazu, analýzu velkých objemů dat, neuronové sítě a integraci s externími službami.

Jinými slovy, představte si, že máte volný a otevřený přístup ke stovkám tisíc hotových softwarových řešení - knihoven a nástrojů, které můžete přímo začlenit do svých podnikových procesů. Je to jako obrovský katalog aplikací pro automatizaci, analýzu, vizualizaci, integraci a další - a všechny jsou k dispozici okamžitě po instalaci Pythonu.

Pandas je jedním z nejoblíbenějších balíčků v ekosystému Pythonu. V roce 2022 dosáhl průměrný počet stažení knihovny Pandas 4 milionů denně (obrázek 3.4-3), zatímco na začátku roku 2025 se toto číslo zvýšilo na 12 milionů stažení denně, což odraží její rostoucí popularitu a široké využití v datové analytice a LLM chatu [86]



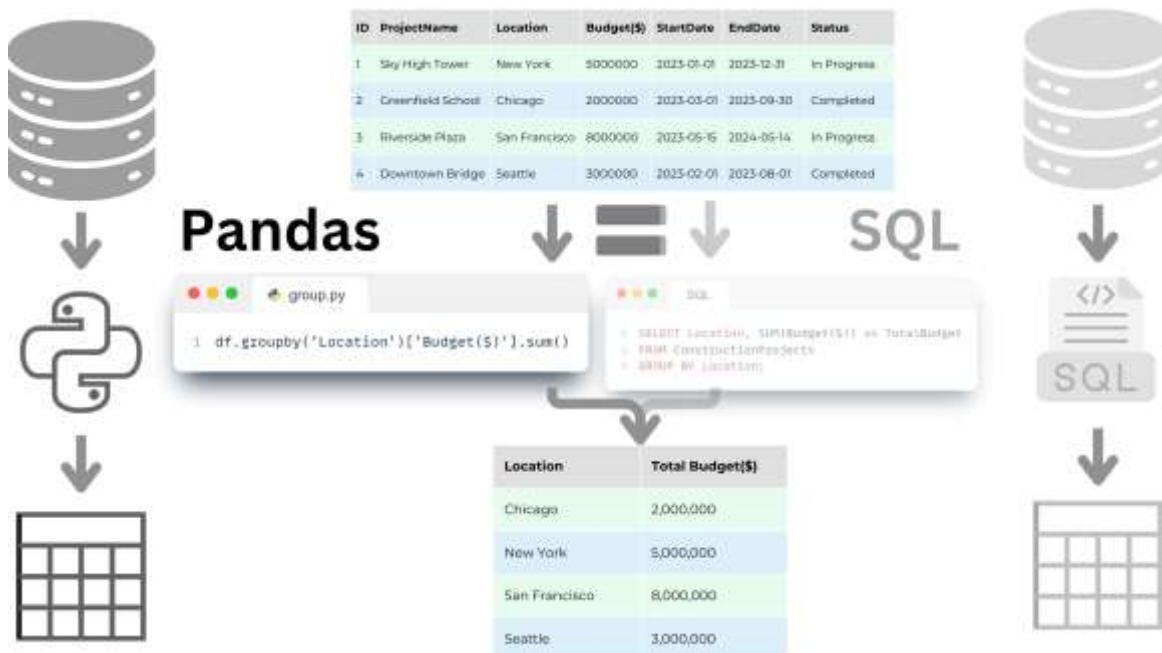
Obrázek 3.4-3 Pandas je jednou z nejstahovanějších knihoven. V roce 2024 přesáhl její roční počet stažení 1,4 miliardy.

Dotazovací jazyk v knihovně Pandas je svou funkčností podobný dotazovacímu jazyku SQL, který jsme probírali v kapitole "Relační databáze a dotazovací jazyk SQL".

Ve světě analytiky a správy strukturovaných dat vyniká Pandas svou jednoduchostí, rychlosťí a výkonem a poskytuje uživatelům širokou škálu nástrojů pro efektivní analýzu a zpracování informací.

Oba nástroje - SQL a Pandas - poskytují výkonné možnosti manipulace s daty, zejména ve srovnání s

tradičním Excelom. Podporuje operace, jako je výběr, filtrování (obrázek 3.4-4), s tím rozdílem, že SQL je optimalizován pro práci s relačními databázemi, zatímco Pandas zpracovává data v paměti RAM, což mu umožňuje běžet na jakémkoli počítači, bez nutnosti vytvářet databáze a nasazovat samostatnou infrastrukturu.

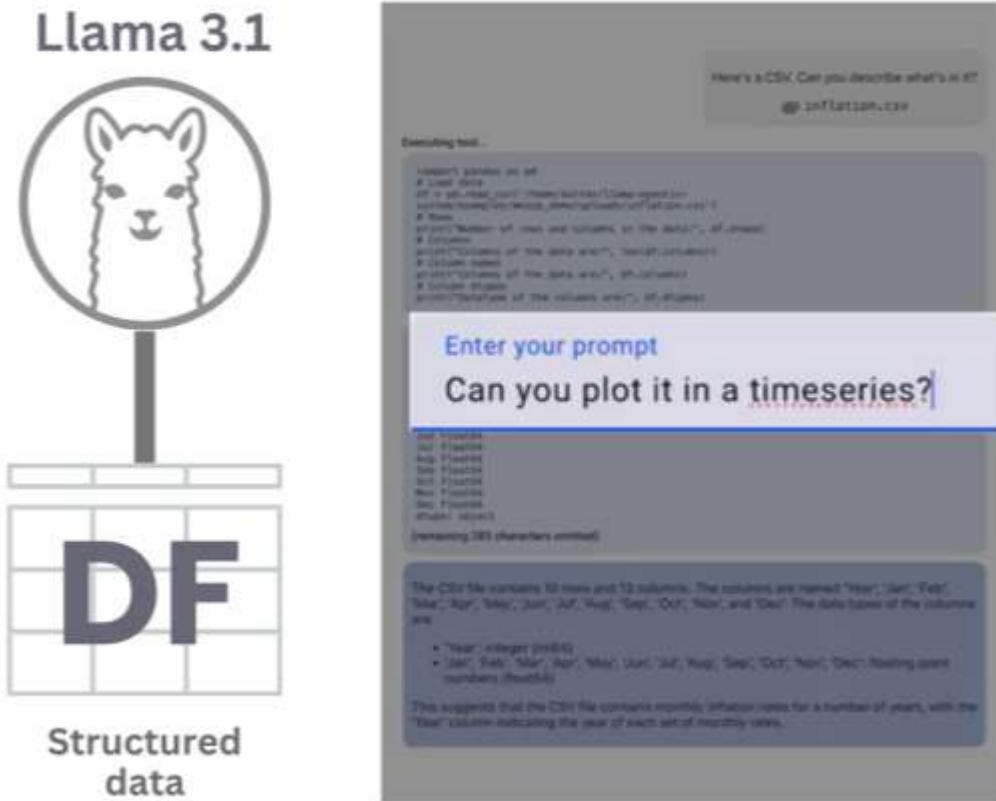


Obr. 3.4-4 Pandas, na rozdíl od SQL, má flexibilitu pro práci s různými datovými formáty, neomezuje se pouze na databáze.

Pandas je často upřednostňován pro vědecký výzkum, automatizaci procesů, vytváření pipeline (včetně ETL) a manipulaci s daty v Pythonu, zatímco SQL je standardem pro správu databází a často se používá v podnikovém prostředí pro zpracování velkého množství dat.

Knihovna Pandas programovacího jazyka Python umožňuje provádět nejen základní operace, jako je čtení a zápis tabulek, ale také složitější úlohy, včetně slučování dat, seskupování dat a provádění složitých analytických výpočtů.

Dnes se knihovna Pandas používá nejen v akademickém výzkumu a podnikové analytice, ale také ve spojení s LLM -modely. Například divize Meta® (Facebook™) při vydání nového open source modelu LLaMA 3.1 v roce 2024 věnovala zvláštní pozornost práci se strukturovanými daty, přičemž jedním z klíčových a prvních případů v jeho vydání bylo právě zpracování strukturovaných datových rámců (obr. 3.4-5) ve formátu CSV a integrace s knihovnou Pandas přímo v chatu.



Obr. 3.4-5 Jedním z prvních a hlavních případů týmu Meta, který byl představen v LlaMa 3.1 v roce 2024, bylo vytváření aplikací pomocí Pandas.

Pandas je základním nástrojem pro miliony datových vědců, kteří zpracovávají a připravují data pro generativní AI. Zrychlení programu Pandas s nulovými změnami kódu bude obrovským krokem vpřed. Datoví vědci budou moci zpracovávat data v řádu minut namísto hodin a získají řádově více dat pro trénování generativních modelů AI [88].

- Jensen Huang, zakladatel a generální ředitel společnosti NVIDIA

Pomocí aplikace Pandas, je možné spravovat a analyzovat datové soubory daleko za hranicemi možností aplikace Excel. Zatímco Excel je obvykle schopen zpracovat až 1 milion řádků dat, Pandas si snadno poradí s datovými sadami (obrázek 9.1-2, obrázek 9.1-10) obsahujícími desítky milionů řádků [89]. Tato schopnost umožňuje uživatelům provádět sofistikovanou analýzu a vizualizaci dat na velkých souborech dat, což poskytuje hluboký vhled a usnadňuje rozhodování založené na datech. Pandas má navíc silnou podporu komunity [90]: stovky milionů vývojářů a analytiků po celém světě (Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker) jej používají online nebo offline každý den, což poskytuje velké množství hotových řešení pro jakýkoli obchodní problém.

Základem většiny analytických procesů v jazyce Python je strukturovaná forma dat zvaná DataFrame,

kterou poskytuje knihovna Pandas. Jedná se o výkonný a flexibilní nástroj pro organizaci, analýzu a vizualizaci tabulkových dat.

DataFrame: univerzální formát tabulkových dat

DataFrame je ústřední struktura knihovny Pandas, což je dvourozměrná tabulka (obrázek 3.4-6), kde řádky odpovídají jednotlivým objektům nebo záznamům a sloupce jejich vlastnostem, parametry nebo kategoriím. Tato struktura vizuálně připomíná tabulky Excelu, ale je mnohem lepší z hlediska flexibility, škálovatelnosti a funkčnosti.

DataFrame je způsob reprezentace a zpracování tabulkových dat uložených v paměti RAM počítače.

DataFrame je způsob reprezentace a zpracování tabulkových dat uložených v paměti RAM počítače. V tabulce mohou řádky odrážet například prvky stavebního projektu a sloupce jejich vlastnosti: kategorie, rozměry, souřadnice, náklady, termíny atd. Taková tabulka navíc může obsahovat jak informace o jednom projektu (obr. 4.1-13), tak údaje o milionech objektů z tisíců různých projektů (obr. 9.1-10). Díky vektorizovaným operacím Pandas lze takové objemy informací snadno filtrovat, seskupovat a agregovat vysokou rychlostí.

The diagram illustrates a DataFrame as a structured data frame. It features a building crane icon above a table. The table has columns labeled: ID, Name, Category, Family Name, Height, BoundingBoxMin_X, BoundingBoxMin_Y, BoundingBoxMin_Z, and Level. Annotations include:

- Index label:** Points to the first column (ID).
- Index axis=0:** Points to the first row (431144).
- Missing value:** Points to a blank cell in the Height column.
- Column names:** Points to the top row of the table.
- STRUCTURED DATA:** Text next to the table.
- Level:** Column header for the last column.

ID	Name	Category	Family Name	Height	BoundingBoxMin_X	BoundingBoxMin_Y	BoundingBoxMin_Z	Level
431144	Single-Flush	OST_Doors	Single-Flush	6.88976378	20.1503	-10.438	9.84252	Level 1
431198	Single-Flush	OST_Doors		6.88976378	13.2281	-1.1207	9.84252	Level 2
457479	Single Window	OST_Windows	Single Window	8.858267717	-11.434	-11.985	9.80971	Level 2
485432	Single Window	OST_Windows	Single Window	8.858267717	-11.434	4.25986	9.80971	Level 2
490150	Single-Flush	OST_Doors	Single-Flush	6.88976378	-1.5748	-2.9565	-1E-16	Level 1
493697	Basic Wall	OST_Walls	Basic Wall		-38.15	20.1656	-4.9213	Level 1
497540	Basic Wall	OST_Walls	Basic Wall		-4.5212	0.0708	9.84252	Level 1

Obrázek 3.4-6 Konstrukční projekt jako datový rámec je dvourozměrná tabulka s prvky v řadách a atributy ve sloupcích.

Společnost Nvidia odhaduje, že již dnes se až 30% všech výpočetních zdrojů využívá ke zpracování strukturovaných dat - datových rámců - a tento podíl stále roste.

V každé firmě se zpracovává pravděpodobně třetina světové výpočetní techniky.
Zpracování dat a data většiny firem jsou v DataFrame, ve formátu tabulky

- Jensen Huang, generální ředitel společnosti Nvidia [91]

Uvedeme si některé klíčové funkce DataFrame v Pandas:

- Sloupce: v DataFrame jsou data uspořádána do sloupců, z nichž každý má jedinečný název. Sloupce atributů mohou obsahovat data různých typů, podobně jako sloupce v databázích nebo sloupce v tabulkách.
- **Pandas Řada** je v programu Pandas jednorozměrná datová struktura podobná seznamu nebo sloupci v tabulce, kde každá hodnota odpovídá jinému indexu.

Řada Pandas má více než 400 atributů a metod, což umožňuje neuvěřitelně flexibilní práci s daty. Na sloupec můžete přímo aplikovat jednu ze čtyř set dostupných funkcí, provádět matematické operace, filtrovat data, nahrazovat hodnoty, pracovat s daty, řetězci a mnoha dalšími funkcemi. Řada Series navíc podporuje vektorové operace, což oproti cyklickým výpočtům výrazně urychluje zpracování velkých datových sad. Můžete například snadno vynásobit všechny hodnoty číslem, nahradit chybějící data nebo použít složité transformace bez nutnosti psát složité smyčky.

- **Řádky**: v DataFrame lze indexovat jedinečnými hodnotami. Tento index umožňuje rychle měnit a upravovat data v konkrétních řádcích.
- **Index**: Ve výchozím nastavení při vytváření DataFrame Pandas přiřadí každému řádku index od 0 do N-1 (kde N je počet všech řádků v DataFrame). Index však lze změnit tak, aby obsahoval speciální označení, jako jsou data nebo jedinečné charakteristiky.
- **Indexování** řádků v DataFrame znamená, že každému řádku je přiřazen jedinečný název nebo označení, které se nazývá index DataFrame.
- **Datové typy**: DataFrame podporuje celou řadu datových typů včetně: `int`, `float`, `bool`, `datetime64` a `object` pro textová data. Každý sloupec DataFrame má svůj vlastní datový typ, který určuje, jaké operace lze s jeho obsahem provádět.
- **Operace s daty**: DataFrame podporuje širokou škálu operací pro zpracování dat, včetně agregace (`groupby`), slučování (`merge` a `join`), spojování (`concat`), split-apply-combine a mnoha dalších technik transformace dat.
- **Manipulace s velikostí**: DataFrame umožňuje přidávat a odebírat sloupce a řádky, takže se jedná o dynamickou strukturu, kterou lze upravovat podle potřeb analýzy dat.
- **Vizualizace dat**: Pomocí vestavěných vizualizačních technik nebo propojením s populárními knihovnami pro vizualizaci dat, jako je Matplotlib nebo Seaborn, lze DataFrame snadno převést na grafy a diagramy a graficky prezentovat data.
- **Vstup a výstup dat**: Pandas poskytuje funkce pro čtení import a export dat do různých formátů souborů, jako jsou CSV, Excel, JSON, HTML a SQL, čímž se DataFrame může stát centrálním centrem pro sběr a distribuci dat.

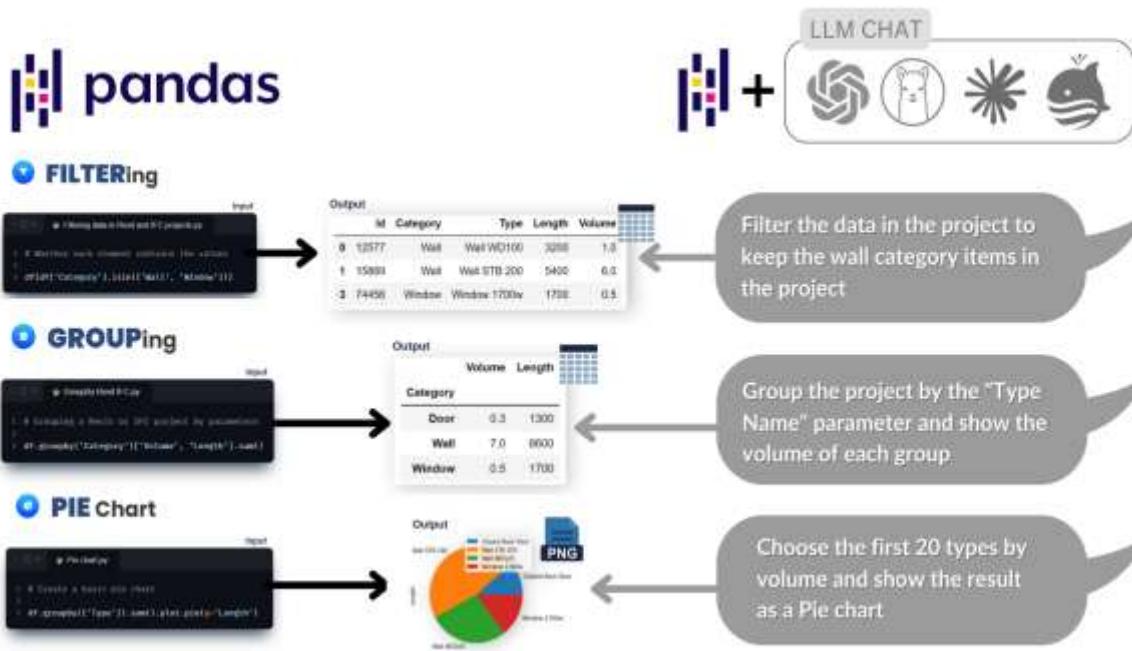
Na rozdíl od CSV a XLSX, Pandas DataFrame poskytuje větší flexibilitu a výkon při práci s daty: dokáže zpracovat velké množství informací v paměti RAM, podporuje rozšířené datové typy (včetně dat, logických hodnot a časových řad) a poskytuje rozsáhlé možnosti filtrování, agregace, slučování a

vizualizace dat. Zatímco CSV neukládá informace o typech a struktuře dat a XLSX je často přetížený formátováním a má nízkou škálovatelnost, DataFrame zůstává optimální volbou pro rychlou analýzu, automatizaci procesů a integraci s AI -modely (obrázek 3.4-7). Následující kapitoly se budou podrobně zabývat každým z těchto aspektů dat, v 8. části knihy budou také podrobně probrány podobné formáty, jako jsou Parquet, Apache Orc, JSON, Feather, HDF5 a datové sklady (obr. 8.1-2).

		XLSX	CSV	Pandas DataFrame
	Storage	Tabular	Tabular	Tabular
	Usage	Office tasks, data presentation	Simple data exchange	Data analysis, manipulation
	Compression	Built-in	None	None (in-memory)
	Performance	Low	Medium	High (memory dependent)
	Complexity	High (formatting, styles)	Low	Low
	Data Type Support	Limited	Very limited	Extended
	Scalability	Low	Low	Medium (memory limited)

Obr. 3.4-7 DataFrame je optimální volbou pro manipulaci s daty s vysokým výkonem a pokročilou podporou datových typů.

Díky své flexibilitě, výkonu a snadnému použití se knihovna Pandas a formát DataFrame staly de facto standardem v analýze dat v jazyce Python. Jsou ideální pro vytváření jednoduchých sestav i pro budování složitých analytických pipeline, zejména ve spojení s modely LLM.



Obr. 3.4-8 LLM zjednodušuje interakci s Pandas: místo kódu stačí textový dotaz.

Dnes se Pandas aktivně používá v chatovacích místnostech založených na LLM, jako jsou ChatGPT, LLaMa, DeepSeek, QWEN a další. V mnoha případech, kdy model obdrží dotaz týkající se zpracování tabulek, validace dat nebo analýzy, generuje kód právě pomocí knihovny Pandas. Díky tomu je DataFrame přirozeným "jazykem" pro reprezentaci dat v dialozích s umělou inteligencí (obrázek 3.4-8).

Moderní datové technologie, jako je Pandas, usnadňují analýzu, automatizaci a integraci dat do podnikových procesů. Poskytují rychlé výsledky, snižují pracovní zátěž specialistů a zajišťují opakovatelnost operací.

Další kroky: budování udržitelného datového rámce

V této části jsme se věnovali hlavním typům dat používaných ve stavebnictví, seznámili se s různými formáty pro jejich ukládání a analyzovali roli moderních nástrojů, včetně LLM a IDE, při zpracování informací. Dozvěděli jsme se, že efektivní správa dat je základem pro informované rozhodování a automatizaci podnikových procesů. Organizace, které jsou schopny strukturovat a organizovat svá data, získávají významnou konkurenční výhodu ve fázích zpracování a transformace dat.

Pro shrnutí této části je vhodné zdůraznit hlavní praktické kroky, které vám pomohou aplikovat diskutované přístupy ve vašich každodenních úkolech:

- Proveďte datový audit svých procesů
- Proveďte soupis všech typů dat, které používáte ve svých projektech.

- Určete, které typy dat a modely jsou pro vaše obchodní procesy nejdůležitější.
- Identifikovat problémové oblasti, kde informace zůstávají často nestrukturované, špatně strukturované nebo nedostupné.

■ Začněte vytvářet strategii správy dat

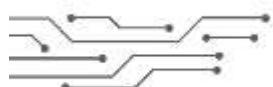
- Zvyšování politických otázek a standardů pro práci s různými typy dat.
- Analyzujte, které z vašich pracovních postupů lze zlepšit převedením nestrukturovaných dat na strukturovaná data.
- Vytvoření zásad ukládání dat a přístupu k nim, které zohledňují bezpečnost a důvěrnost.

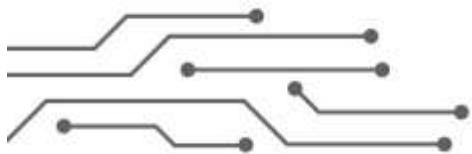
■ Instalace a ovládání základních nástrojů pro práci s daty

- Vyberte si vhodné IDE, které odpovídá vašim úkolům (např. nainstalujte VS Code nebo Jupyter Notebook).
- Zkuste si založit místní LLM, která bude důvěrně zpracovávat vaše osobní údaje.
- Začněte experimentovat s knihovnou Pandas a zpracovávejte tabulková data XLSX.
- Popište studentovi LLM typické úlohy, které řešíte v tabulkových nástrojích nebo databázích, a požádejte ho o automatizaci práce pomocí programu Pandas.

Použití těchto kroků vám umožní postupně změnit přístup k práci s daty a přejít od nesourodých, nestrukturovaných datových souborů k jednotnému ekosystému, kde se data stanou dostupným a srozumitelným přínosem. Začněte v malém - vytvořte svůj první DataFrame v programu Pandas, spusťte lokální LLM, zautomatizujte svou první rutinní úlohu pomocí Pythonu (např. tabulky v Excelu).

Čtvrtá část knihy se zaměří na kvalitu dat, organizaci, strukturování a modelování dat. Zaměříme se na metodiky, které transformují různorodé zdroje dat - od souborů PDF a textů po obrázky a modely CAD - do strukturovaných datových souborů vhodných pro analýzu a automatizaci. Budeme také zkoumat, jak se formalizují požadavky na data, jak se vytvářejí konceptuální a logické modely v konstrukčních projektech a jak mohou v tomto procesu pomoci moderní jazykové modely (LLM).





IV ČÁST

KVALITA DAT: ORGANIZACE, STRUKTUROVÁNÍ, MODELOVÁNÍ.

Čtvrtá část se zaměřuje na metodiky a technologie pro transformaci různorodých informací do kvalitních strukturovaných datových souborů.

Podrobně jsou rozebrány procesy tvorby a dokumentace datových požadavků jako základu pro efektivní informační architekturu ve stavebních projektech.

Jsou uvedeny praktické metody získávání strukturovaných informací z různých zdrojů (PDF -dokumenty, obrázky, textové soubory, CAD -modely) s příklady realizace. Je rozebráno použití regulárních výrazů (RegEx) a dalších nástrojů pro automatickou validaci a verifikaci dat. Postupně je popsán proces modelování dat na koncepční, logické a fyzické úrovni s přihlédnutím ke specifikům stavebnictví. Jsou demonstrovány konkrétní příklady využití jazykových modelů (LLM) k automatizaci procesů strukturování a ověřování informací. Jsou navrženy efektivní přístupy k vizualizaci výsledků analýzy, které zvyšují dostupnost analytických informací pro všechny úrovně řízení stavebních projektů

KAPITOLA 4.1.

PŘEVOD DAT DO STRUKTUROVANÉ PODOBY

V éře ekonomiky založené na datech se data stávají spíše základem pro rozhodování než překážkou. Namísto neustálého přizpůsobování informací každému novému systému a jeho formátům se společnosti stále více snaží vytvořit jednotný strukturovaný datový model, který slouží jako univerzální zdroj pravdy pro všechny procesy. Moderní informační systémy nejsou navrhovány na základě formátů a rozhraní, ale na základě významu dat - protože struktura se může měnit, ale význam informací zůstává mnohem déle stejný.

Klíčem k efektivní práci s daty není jejich nekonečná konverze a transformace, ale jejich správná organizace od samého počátku: vytvoření univerzální struktury schopné zajistit transparentnost, automatizaci a integraci ve všech fázích životního cyklu projektu.

Tradiční přístup nutí k ručním úpravám při každé implementaci nové platformy: migraci dat, změně názvů atributů, úpravě formátů. Tyto kroky nezlepšují kvalitu samotných dat, ale pouze maskují problémy a vytvářejí začarovaný kruh nekonečných transformací. V důsledku toho se společnosti stávají závislé na konkrétních softwarových řešeních a digitální transformace se zpomaluje.

V následujících kapitolách se podíváme na to, jak správně strukturovat data a jak vytvořit univerzální modely, minimalizovat závislost na platformě a zaměřit se na to nejdůležitější - data jako strategický zdroj, kolem kterého se budují udržitelné procesy.

Naučit se převádět dokumenty, PDF, obrázky a texty do strukturovaných formátů.

Ve stavebních projektech existuje naprostá většina informací v nestrukturované podobě: technické dokumenty, výkazy práce, výkresy, specifikace, harmonogramy, protokoly. Jejich různorodost - co do formátu i obsahu - ztěžuje integraci a automatizaci.

Proces převodu do strukturovaných nebo polostrukturovaných formátů se může lišit v závislosti na typu vstupních dat a požadovaných výsledcích zpracování.

Transformace dat z nestrukturované do strukturované podoby je umění i věda. Tento proces se liší v závislosti na typu vstupních dat a účelu analýzy a často zabírá významnou část práce datového inženýra (obrázek 3.2-5) a analytika, přičemž cílem je vytvořit čistý, uspořádaný soubor dat.



Obr. 4.1-1 Převod nestrukturovaného naskenovaného dokumentu do formátu strukturované tabulky.

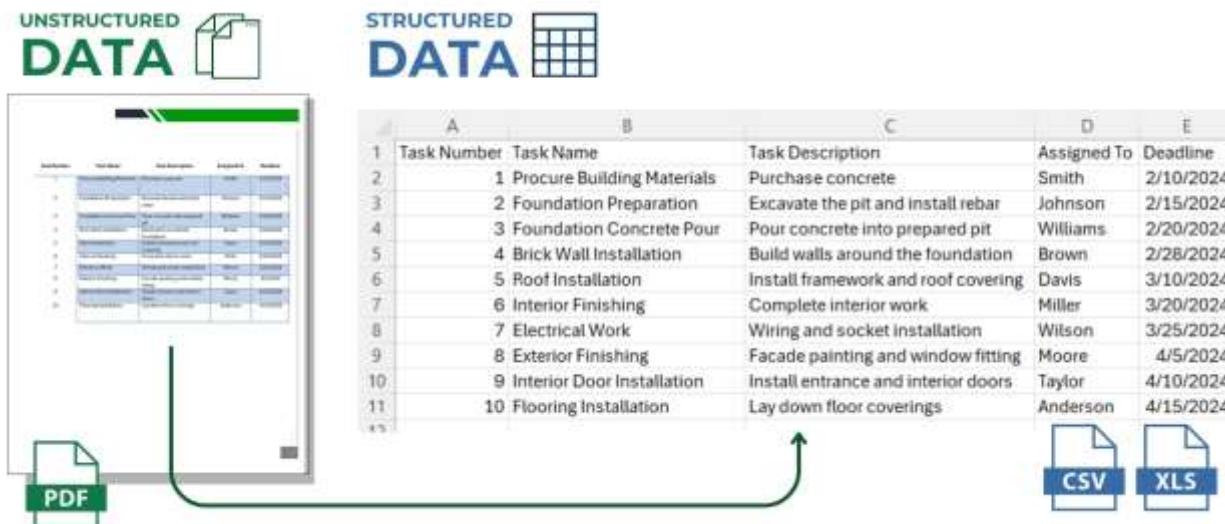
Převod dokumentů, souborů PDF, obrázků a textů do strukturovaného formátu (obrázek 4.1-1) je postupný proces, který zahrnuje následující kroky:

- **Výpis**): V tomto kroku se načte zdrojový dokument nebo obrázek obsahující nestrukturovaná data. Může to být například dokument PDF -dokument, fotografie, výkres nebo schéma.
- **Transformace dat** (Transform): Následuje krok transformace nestrukturovaných dat do strukturovaného formátu. Může jít například o rozpoznání a interpretaci textu z obrázků pomocí optického rozpoznávání znaků (OCR) nebo jiných metod zpracování.
- **Načtení a uložení dat** (Load): Poslední krok zahrnuje uložení zpracovaných dat v různých formátech, například CSV, XLSX, XML, JSON, pro další práci, přičemž volba formátu závisí na konkrétních požadavcích a preferencích.

Tento proces, známý jako ETL (Extract, Transform, Load), hraje klíčovou roli při automatizovaném zpracování dat a bude podrobněji popsán v kapitole "ETL a Pipeline: Extract, Transform, Load". Dále se podíváme na příklady transformace dokumentů různých formátů na strukturovaná data.

Příklad převodu dokumentu PDF na tabulku

Jedním z nejčastějších úkolů ve stavebních projektech je zpracování specifikací ve formátu PDF. Pro demonstraci přechodu z nestrukturovaných dat do strukturovaného formátu si uvedeme praktický příklad: extrakce tabulky z dokumentu PDF a její převod do formátu CSV nebo Excel (obr. 4.1-2).



Obr. 4.1-2 Na rozdíl od PDF jsou formáty CSV a XLSX široce rozšířené a snadno se integrují do různých systémů pro správu dat.

Jazykové modely LLM, například ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN výrazně zjednodušují způsob, jakým datoví vědci pracují s daty, snižují potřebu hlubokého učení programovacích jazyků a umožňují řešit mnoho úloh pomocí textových dotazů.

Proto místo toho, abychom trávili čas hledáním řešení na internetu (obvykle na webu StackOverflow nebo tematických fórech a chatech) nebo kontaktováním specialistů na zpracování dat, můžeme využít možností moderních online nebo lokálních LLM. Stačí položit dotaz a model poskytne hotový kód pro převod PDF -dokumentu do tabulkového formátu.

- ✍ Pošlete následující textový požadavek libovolnému modelu LLM (CHATGRT, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN nebo jinému):

Napište prosím kód pro extrakci textu ze souboru PDF, který obsahuje tabulku. Kód by měl jako argument přijmout cestu k souboru a vrátit extraovanou tabulku jako DataFrame ↴

- 💡 Odpověď LLM -modelu bude ve většině případů ve formě kódu v jazyce Python , protože tento jazyk je široce používán pro zpracování dat, automatizaci a práci s různými formáty souborů:

```

1 import fitz # PyMuPDF
2 import pandas as pd
3 import re
4
5 # Define the path to the PDF file
6 pdf_path = r'C\Task Number.pdf'
7
8 # Step 1: Extracting text from the PDF
9 def extract_text(pdf_path):
10     """
11     Extract text from a PDF file.
12     Args: pdf_path (str): Path to the PDF file.
13     Returns: str: All text extracted from the PDF.
14     """
15     document = fitz.open(pdf_path)
16     text = ''
17     for page in document:
18         text += page.get_text()
19     return text
20
21 extracted_text = extract_text(pdf_path)

```

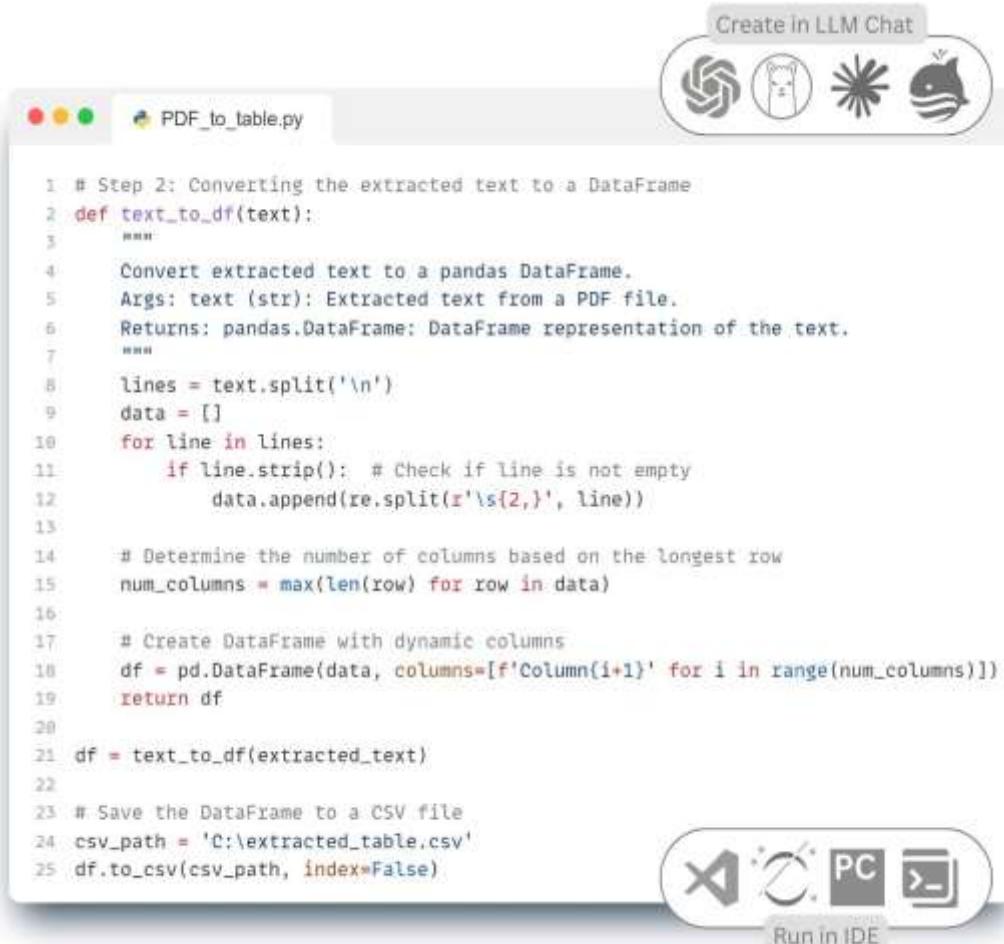
Obr. 4.1-3 Odezva LLM v podobě kódu jazyka Python a jeho knihoven a balíčků (Pandas, Fitz) extrahuje text z PDF -souboru.

Tento kód (obrázek 4.1-3) lze spustit offline v některém z populárních IDE, které jsme zmínili výše: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse s pluginem PyDev, Thonny, Wing IDE, IntelliJ IDEA s pluginem Python, JupyterLab nebo populární online nástroje: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- 💡 V kroku "Convert" použijeme populární knihovnu Pandas (o které jsme podrobně hovořili v kapitole "Python Pandas: nepostradatelný nástroj pro práci s daty") k načtení extrahovaného textu do DataFrame a uložení DataFrame do souboru tabulky CSV nebo XLSX:

Potřebuji kód, který převede výslednou tabulku ze souboru PDF -file do DataFrame.
Také přidat kód, který uloží DataFrame do souboru CSV. ↴

2 Odpověď společnosti LLM :



The screenshot shows a code editor window titled "PDF_to_table.py". The code is as follows:

```

1 # Step 2: Converting the extracted text to a DataFrame
2 def text_to_df(text):
3     """
4         Convert extracted text to a pandas DataFrame.
5         Args: text (str): Extracted text from a PDF file.
6         Returns: pandas.DataFrame: DataFrame representation of the text.
7     """
8     lines = text.split('\n')
9     data = []
10    for line in lines:
11        if line.strip(): # Check if line is not empty
12            data.append(re.split(r'\s{2,}', line))
13
14    # Determine the number of columns based on the longest row
15    num_columns = max(len(row) for row in data)
16
17    # Create DataFrame with dynamic columns
18    df = pd.DataFrame(data, columns=[f'Column{i+1}' for i in range(num_columns)])
19    return df
20
21 df = text_to_df(extracted_text)
22
23 # Save the DataFrame to a CSV file
24 csv_path = 'C:\extracted_table.csv'
25 df.to_csv(csv_path, index=False)

```

At the top right of the window are icons for "Create in LLM Chat" and four AI models. At the bottom right are icons for "Run in IDE" and "PC".

Obr. 4.1-4 Převedení extrahované tabulky z PDF do DataFrame a uložení tabulky do souboru CSV .

Pokud se během provádění kódu vyskytne chyba (obr. 4.1-3, obr. 4.1-4) - např. kvůli chybějícím knihovnám nebo špatné cestě k souboru - lze text chyby jednoduše zkopirovat spolu se zdrojovým kódem a znova odeslat do LLM -modelu. Model analyzuje chybové hlášení, vysvětlí, v čem je problém, a navrhne opravy nebo další kroky.

Interakce s AI LLM se tak stává uceleným cyklem →→test→ zpětná vazba→ oprava - bez nutnosti hlubokých technických znalostí.

Pomocí prostého textového dotazu v LLM chat a tuctu řádků jazyka Python, který můžeme spustit lokálně v libovolném IDE, jsme převedli dokument PDF -dokument do tabulkového formátu CSV, který je na rozdíl od dokumentu PDF snadno strojově čitelný a rychle integrovatelný do libovolného systému pro správu dat.

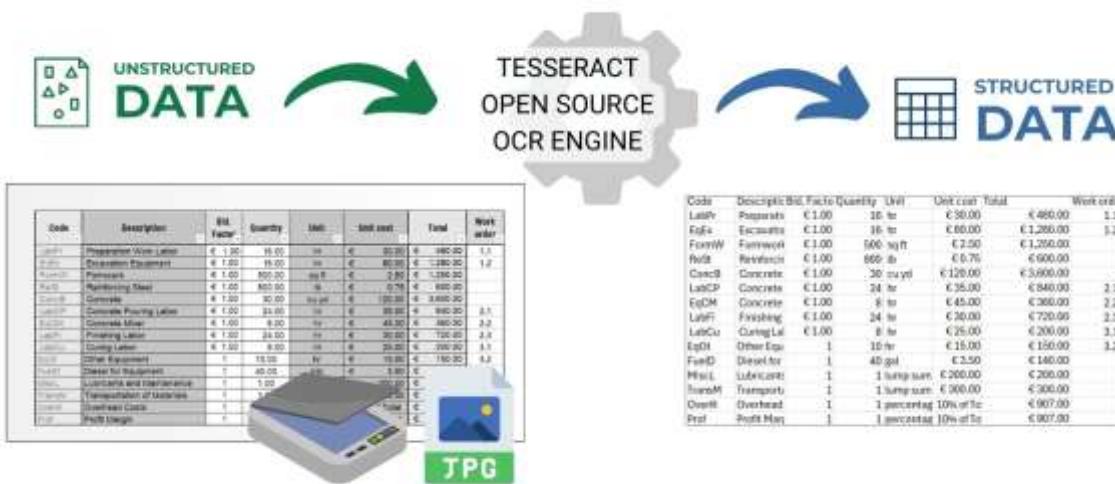
Tento kód (obrázek 4.1-3, obrázek 4.1-4) můžeme zkopirováním z libovolného chatu LLM použít desítky nebo tisíce nových dokumentů PDF na serveru, čímž automatizujeme proces převodu proudu nestrukturovaných dokumentů do strukturovaného formátu tabulky CSV.

Dokumenty PDF však ne vždy obsahují text, častěji se jedná o naskenované dokumenty, které je třeba zpracovat jako obrázky. Ačkoli jsou obrázky ze své podstaty nestrukturované, vývoj a použití knihoven pro rozpoznávání nám umožňuje extrahovat, zpracovávat a analyzovat jejich obsah, což nám umožňuje plně využívat tato data v našich obchodních procesech.

Převod obrázku JPEG, PNG do strukturované podoby

Obrázky jsou jednou z nejběžnějších forem nestrukturovaných dat. Ve stavebnictví a v mnoha dalších odvětvích je obrovské množství informací uloženo ve formě naskenovaných dokumentů, schémat, fotografií a výkresů. Taková data obsahují cenné informace, ale nelze je přímo zpracovat, například v tabulkovém procesoru Excel nebo v databázi. Obrázky obsahují mnoho složitých informací, protože jejich obsah, barvy, textury jsou různorodé a k získání užitečných informací je nutné speciální zpracování.

Potíž při používání obrázků jako zdroje dat spočívá v nedostatečné strukturovanosti. Obrázky nepředávají význam přímým, snadno kvantifikovatelným způsobem, který by počítač mohl okamžitě pochopit nebo zpracovat, jako to dělá tabulka v Excelu nebo tabulka v databázi. K převodu nestrukturovaných obrazových dat do strukturované podoby je třeba použít speciální knihovny schopné interpretovat vizuální informace, které obsahují (obrázek 4.1-5).



Obr. 4.1-5 Převod naskenovaných dokumentů a obrázků do strukturovaných formátů je možný pomocí speciálních nástrojů OCR.

Technologie OCR (Optické rozpoznávání znaků) se používá k extrakci textu z obrázků. Umožňuje rozpoznávat písmena a číslice ve skenovaných dokumentech, fotografiích a souborech PDF a převádět je na upravitelný a strojově čitelný text. Technologie OCR se již dlouho používají při automatizaci dokumentů a dnes je lze snadno integrovat do jakýchkoli obchodních procesů a aplikací Python - aplikací. Jedním z nejpopulárnějších nástrojů OCR je Tesseract, open source, původně vyvinutý společností HP™ a nyní podporovaný společností Google™. Podporuje více než 100 jazyků a má vysokou přesnost rozpoznávání.

Požádejme LLM chat, aby napsal ukázkový kód pro strukturované načtení dat z naskenované nebo vyfotografované tabulky.

- ❷ Pošlete textovou žádost do chatu LLM (ChatGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN nebo jakéhokoli jiného):

Napište kód pro převod obrázku JPEG obsahujícího tabulku do tabulky DataFrame ↴

- 2 V odpovědi LLM se ve většině případů navrhuje použití knihovny Pytesseract k rozpoznávání textu v obrázcích :



```

1 import pytesseract
2 from PIL import Image
3 import pandas as pd
4
5 # Load the image from file
6 image_path = '/mnt/data/ConstructionDDC'
7 image = Image.open(image_path)
8
9 # Use Tesseract to convert the image to text
10 text = pytesseract.image_to_string(image, lang='eng')
11
12 # Split the text into lines and then columns
13 rows = text.split('\n')
14 data = [row.split('\t') for row in rows if row.strip()]
15
16 # Create a DataFrame using pandas
17 # Assuming the first row is the header
18 df = pd.DataFrame(data[1:], columns=data[0])
19
20 # Display the DataFrame
21 df.head() # Show only the first few rows for brevity
22

```

Obr. 4.1-6 Převod textu získaného z tabulky obrázků nebo fotografií do strukturované tabulkové reprezentace.

V tomto příkladu - kód (obrázek 4.1-6) odvozený z LLM, používá knihovnu pytesseract (Tesseract pro Python) k převodu obrázku na text pomocí OCR (optické rozpoznávání znaků) a knihovnu Pandas k převodu tohoto textu do strukturované podoby, tj. do DataFrame.

Proces převodu obvykle zahrnuje předzpracování pro zlepšení kvality obrazu, po kterém se použijí různé algoritmy pro detekci vzorů, extrakci prvků nebo rozpoznávání objektů. Výsledkem je převod nestrukturovaných vizuálních informací na strukturovaná data.

Ačkoli jsou klíčovým zdrojem nestrukturovaných informací soubory PDF a obrázky, skutečným šampionem z hlediska objemu je text generovaný v e-mailech, chatech, schůzkách a messengerech. Těchto dat je nejen mnoho - jsou rozptýlená, neformalizovaná a velmi špatně strukturovaná.

Převod textových dat do strukturované podoby

Kromě dokumentů PDF s tabulkami (obr. 4.1-2) a naskenovaných verzí tabulkových formulářů (obr. 4.1-5) je významná část informací v projektové dokumentaci prezentována v textové podobě. Mohou to být jak souvislé věty v textových dokumentech, tak útržkovité záznamy roztroušené po výkresech a schématech. V podmínkách moderního zpracování dat je jedním z nejčastějších úkolů převést takový text do strukturovaného formátu vhodného pro analýzu, vizualizaci a rozhodování.

Ústředním prvkem tohoto procesu je taxonomie, klasifikační systém, který umožňuje uspořádat informace do kategorií a podkategorií na základě společných znaků.

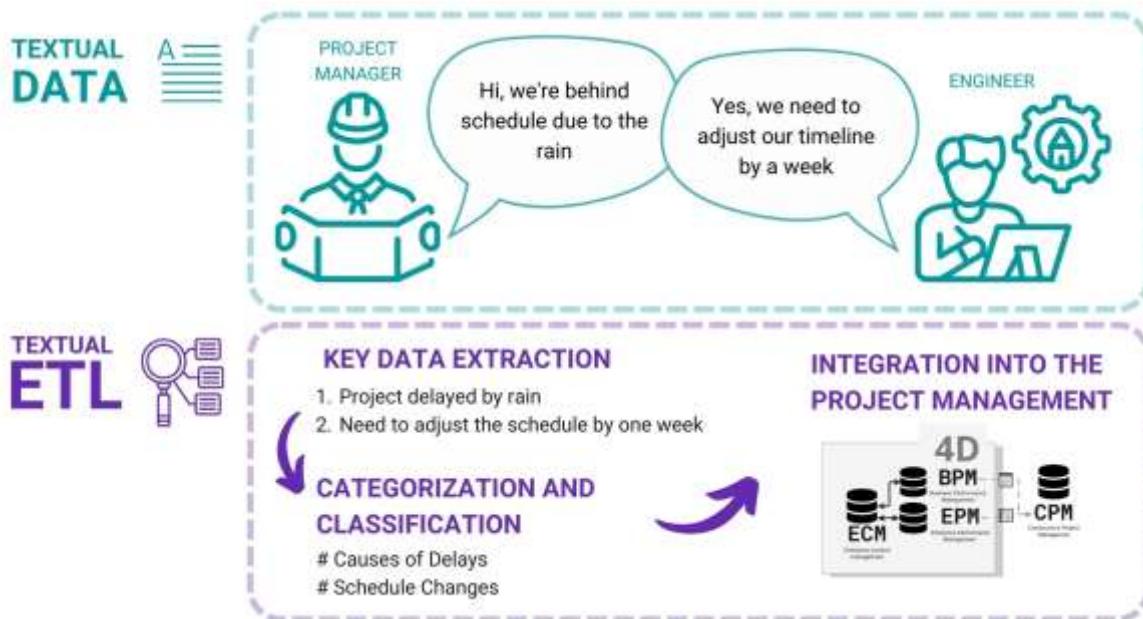
Taxonomie je hierarchická klasifikační struktura, která slouží k seskupování a organizaci objektů. V kontextu zpracování textů slouží jako základ pro systematické třídění položek do sémantických kategorií, čímž zjednodušuje analýzu a zvyšuje kvalitu zpracování dat.

Vytváření taxonomie je doprovázeno kroky extrakce entit, kategorizace a kontextualizace. Pro modelování procesu extrakce informací z textových dat jsou následující kroky podobné těm, které jsme již použili při strukturování dat z PDF dokumentů:

- **Výpis**): je třeba analyzovat textová data a získat informace o zpožděních a změnách v harmonogramu projektu.
- **Kategorizace a klasifikace** (Transofrm): kategorizace obdržených informací, např. důvodů zpoždění a změn v harmonogramu.
- **Integrace** (Load): na závěr připravíme strukturovaná data pro integraci do externích systémů pro správu dat.

Vezměme si situaci: Vedeme dialog mezi projektovým manažerem a inženýrem, který diskutuje o problémech s harmonogramem. Naším cílem je vybrat klíčové prvky (důvody zpoždění, úpravy harmonogramu) a strukturovaně je prezentovat (obrázek 4.1-7).

Provedme extrakci na základě očekávaných klíčových slov, vytvořme DataFrame pro simulaci extrakce dat a po transformaci novou tabulkou DataFrame, která bude obsahovat sloupce pro datum, událost (např. důvod zpoždění) a akci (např. změna plánu).



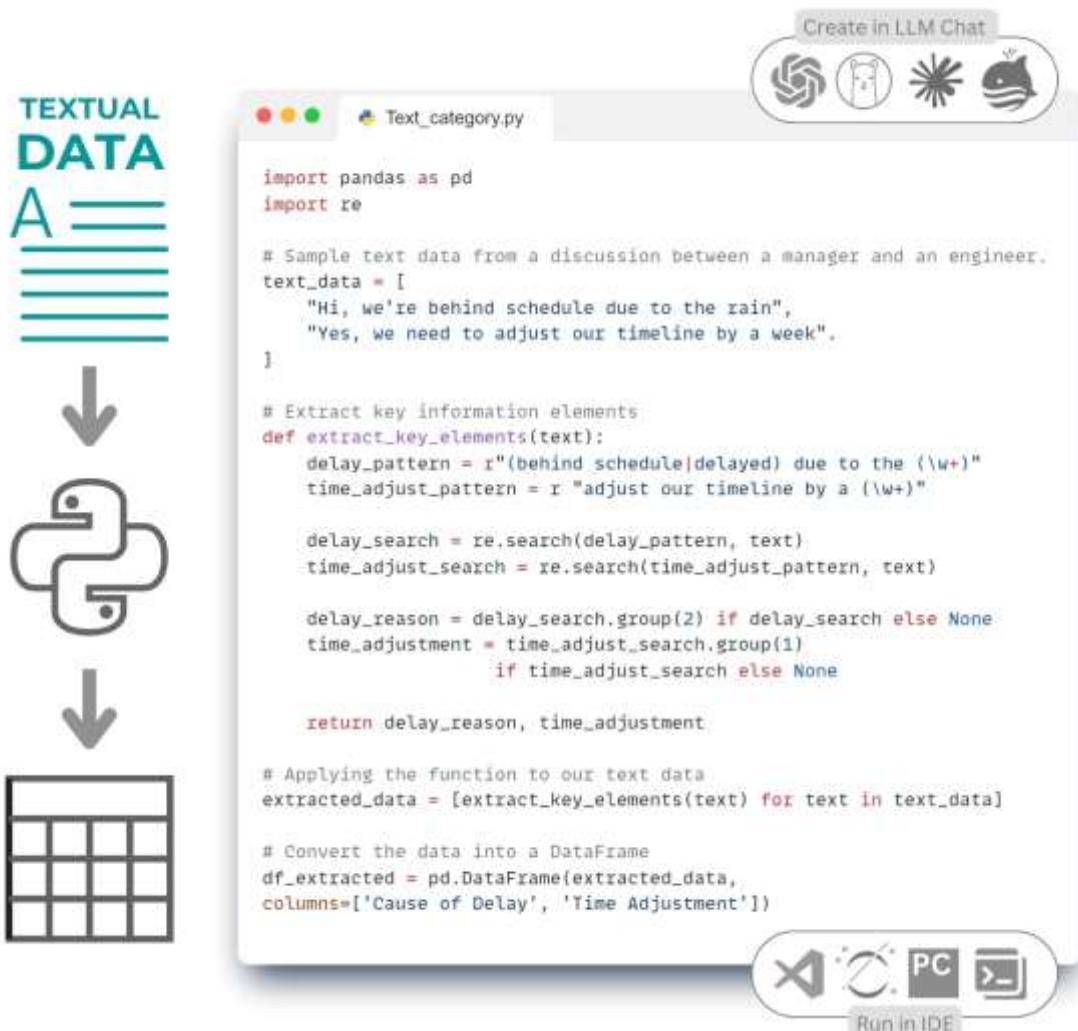
Obrázek 4.1-7 Zvýraznění klíčových informací z textu o potřebě upravit časový plán a začlenit změny do systému řízení projektu.

Zde je kód pro řešení problému pomocí textového dotazu v jednom z jazykových modelů jako v předchozích příkladech.

- ✉ Pošlete textovou žádost do libovolné chatovací místnosti LLM:

Vedu rozhovor mezi manažerem: "Dobrý den, máme zpoždění kvůli dešti" a inženýrem: "Ano, musíme upravit termín o týden". Potřebuji skript, který bude analyzovat budoucí podobné textové dialogy, extrahovat z nich důvody zpoždění a nutných úprav termínů a poté z těchto dat vygeneruje DataFrame. DataFrame by pak měl být uložen do souboru CSV. ↴

- Odpověď z LLM bude obvykle obsahovat kód jazyka Python -kód používající regulární výrazy (re - Regex) a knihovnu Pandas (pd) :

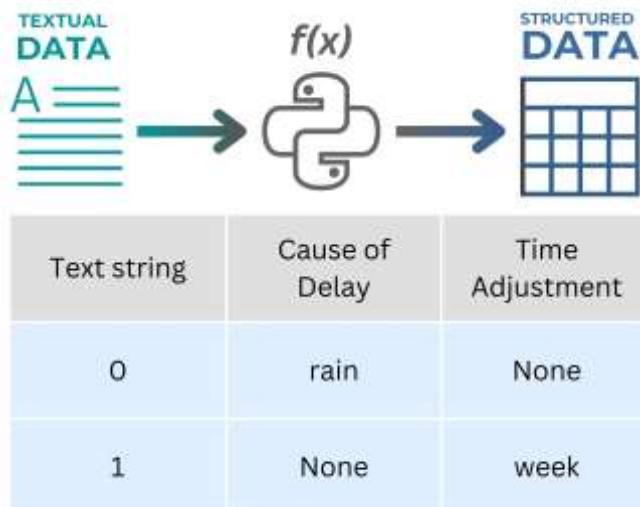


Obrázek 4.1-8 Zvýraznění klíčových informací z textu o potřebě úpravy termínů v tabulce.

V tomto příkladu (obrázek 4.1-7) jsou analyzována textová data obsahující korespondenci mezi projektovým manažerem a inženýrem s cílem identifikovat a extrahovat specifické informace, které mohou ovlivnit řízení budoucích projektů s podobnými dialogy. Pomocí regulárních výrazů (více o regulárních výrazech si povíme v kapitole "Strukturované požadavky a regulární výrazy RegEx") jsou pomocí vzorů identifikovány příčiny zpozdění projektu a potřebné úpravy časového harmonogramu. Funkce zapsaná v tomto příkladu vybírá z řetězců buď příčinu zpozdění, nebo úpravu času na základě vzorů: vybírá slovo za "kvůli" jako příčinu zpozdění nebo slovo za "podle" jako úpravu času.

Pokud je v řádku zmíněno zpozdění způsobené počasím, je jako příčina identifikován "déšť"; pokud je v řádku zmíněna úprava harmonogramu pro určité období, je toto období extrahováno jako časová

úprava (obrázek 4.1-9). Nepřítomnost některého z těchto slov v řádku má za následek hodnotu "Žádné" pro příslušný sloupec atributů.



Obr. 4.1-9 Souhrnná tabulka získaná jako DataFrame, po provedení kódu, obsahuje informace o existenci zpoždění a nutných časových úpravách.

Strukturování a parametrizace podmínek z textu (dialogu, dopisu, dokumentu) umožňuje rychle eliminovat zpoždění na stavbě: například nedostatek pracovníků může ovlivnit tempo prací při špatném počasí, takže firmy, které znají parametry zpoždění z dialogů (obr. 4.1-9) mezi mistrem na stavbě a vedoucím projektu - předem mohou posílit posádku v případě nepříznivé předpovědi.

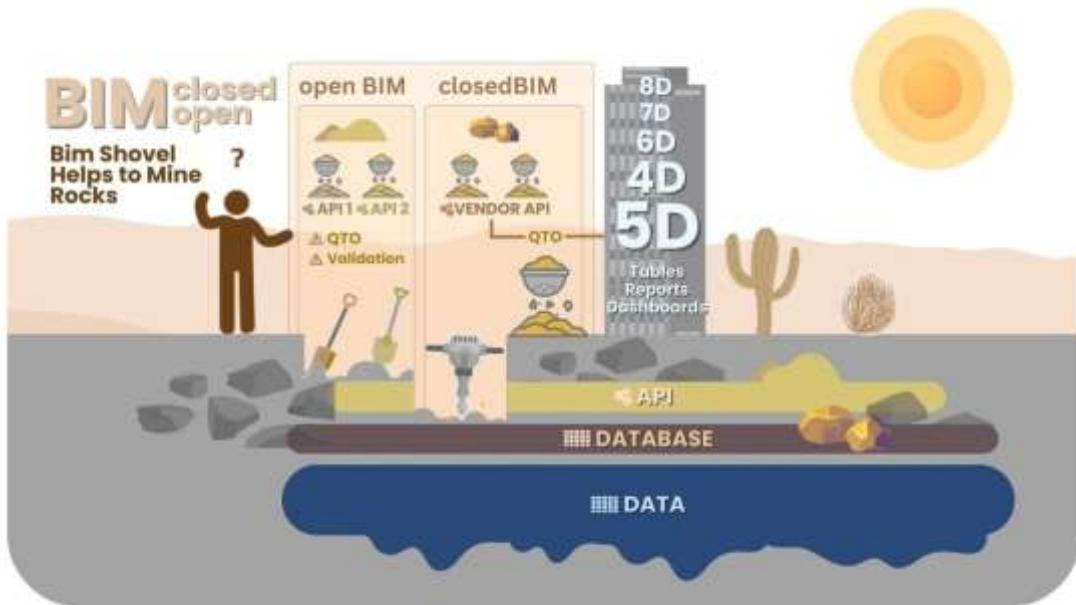
Převod dokumentů a obrázků do strukturovaného formátu lze provést pomocí relativně jednoduchých, otevřených a bezplatných nástrojů založených na kategorizaci.

Klíčovou součástí práce s projektovými daty je také kategorizace prvků, zejména v souvislosti s používáním softwaru CAD (BIM).

Převod dat CAD (BIM) do strukturované podoby.

Strukturování a kategorizace dat CAD (BIM) je složitější úkol, protože data uložená z databází CAD (BIM) jsou téměř vždy v uzavřených nebo složitých parametrických formátech a často kombinují prvky geometrických dat (polostrukturovaná data) a prvky metainformací (polostrukturovaná nebo strukturovaná data) současně.

Nativní datové formáty v systémech CAD (BIM) jsou obvykle chráněny a nepřístupné pro přímé použití, pokud se nejedná o specializovaný software nebo API - rozhraní samotného vývojáře (obr. 4.1-10). Tato izolace dat vytváří uzavřená úložná síla, která omezují volnou výměnu informací a brání vytváření komplexních digitálních procesů ve firmě.



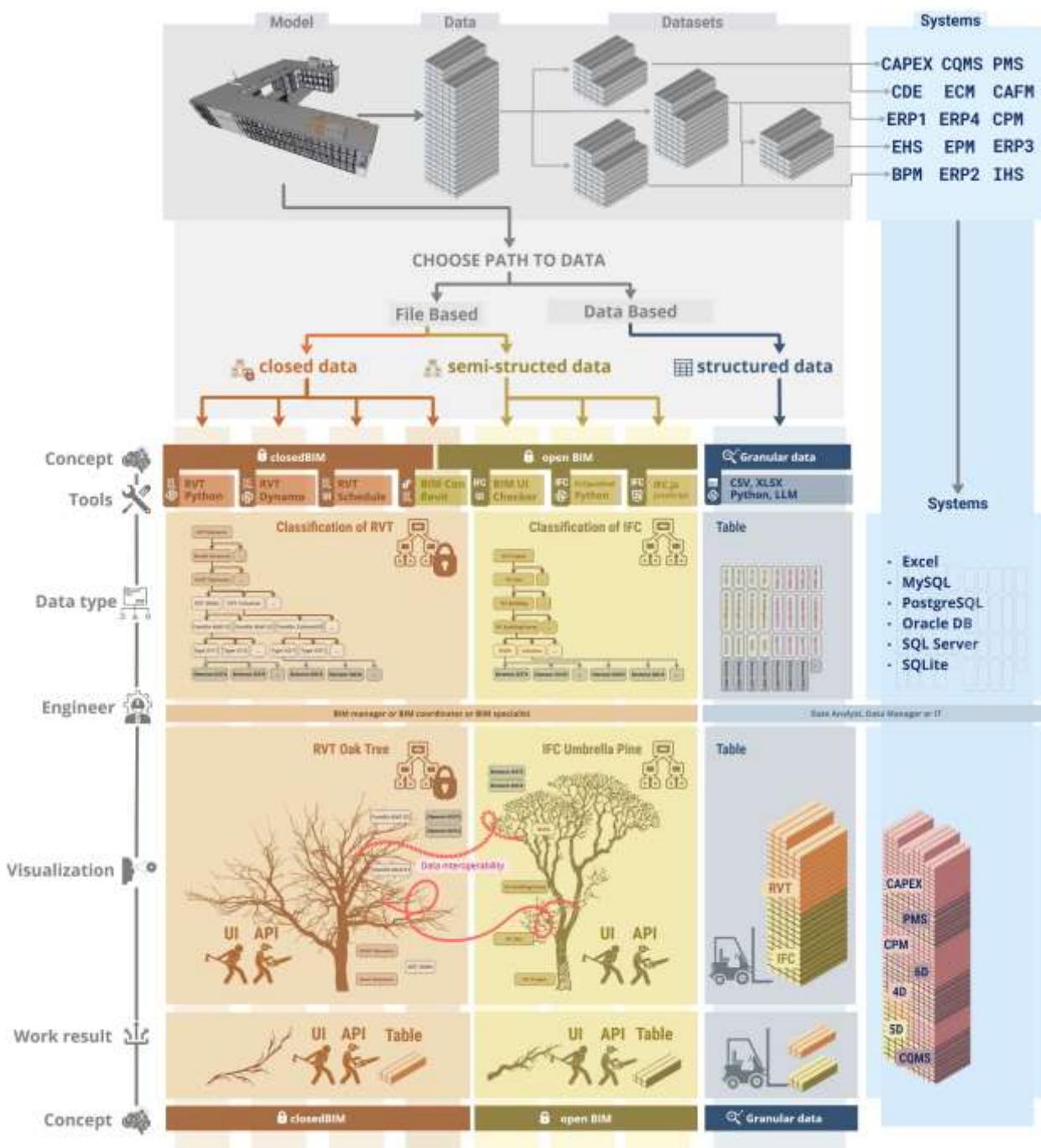
Obrázek 4.1-10 Profesionálové CAD (BIM) mohou přistupovat k nativním datům prostřednictvím API -připojení nebo nástrojů dodavatele.

Ve speciálních formátech CAD (BIM) jsou informace o vlastnostech a atributech prvků projektu shromažďovány v hierarchickém klasifikačním systému, kde jsou entity s odpovídajícími vlastnostmi umístěny, podobně jako plody ovocného stromu, v nejnovějších uzlech větví klasifikace dat (obr. 4.1-11).

Výběr dat z takových hierarchií je možný dvěma způsoby: buď ručně, kliknutím na každý uzel, jako by se zpracovával strom, a osekáním vybraných větví kategorií a typů sekrou. Alternativou je použití aplikacích programových rozhraní (API), které poskytují efektivnější, automatizovaný přístup k získávání a seskupování dat, případně jejich transformaci do strukturované tabulky pro použití v jiných systémech.

K extrakci strukturovaných datových tabulek z projektů CAD (BIM) lze použít různé nástroje, jako jsou Dynamo, pyRvt, Pandamo (Pandas + Dynamo), ACC nebo řešení s otevřeným zdrojovým kódem, například IfcOpSh nebo IFCjs pro formát IFC.

Moderní nástroje pro export a konverzi dat umožňují zjednodušit zpracování a přípravu dat tím, že rozdělují obsah modelů CAD na dvě klíčové složky: geometrické informace a atributová data (obr. 4.1-13) - metainformace popisující vlastnosti konstrukčních prvků (obr. 3.1-16). Tyto dvě datové vrstvy zůstávají propojeny prostřednictvím jedinečných identifikátorů, díky nimž je možné každý prvek s popisem geometrie (prostřednictvím parametrů nebo polygonů) přesně přiřadit k jeho atributům: názvu, materiálu, stupni dokončení, nákladům atd. Tento přístup zajišťuje integritu modelu a umožňuje flexibilní využití dat jak pro vizualizaci (data geometrického modelu), tak pro analytické nebo řídicí úlohy (strukturovaná nebo volně strukturovaná), přičemž s oběma typy dat lze pracovat odděleně nebo souběžně.



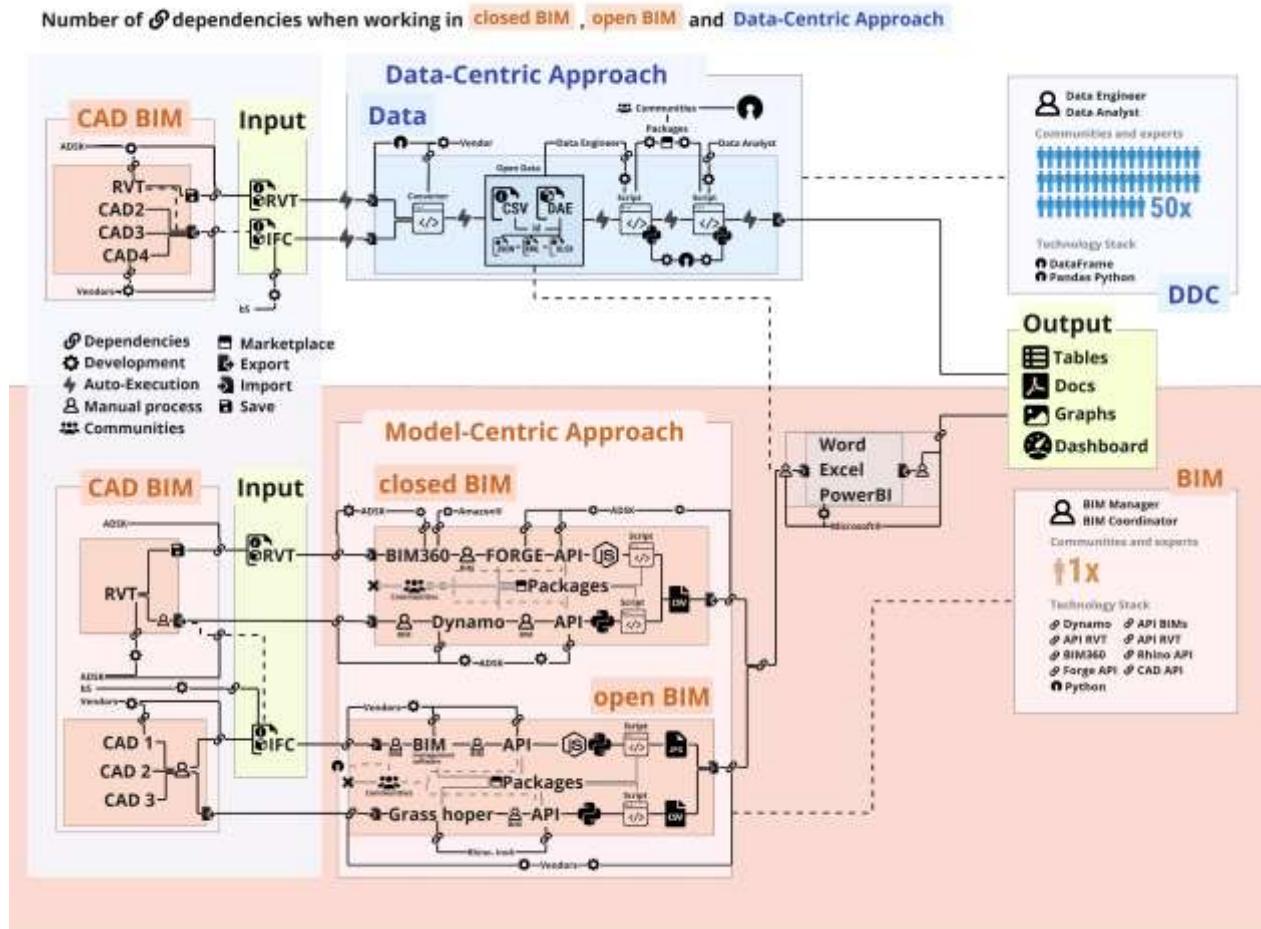
Obr. 4.1-11 Zobrazení informací z databází CAD (BIM) je uživateli prezentováno ve formě klasifikačních stromů.

S rozvojem technologií reverzního inženýrství a příchodem SDK (Software Development Kit) pro konverzi dat CAD se dostupnost a konverze dat z uzavřených formátů softwaru CAD (BIM) stala mnohem snazší. Nyní je možné legálně a bezpečně převádět data z uzavřených formátů do univerzálních formátů vhodných pro analýzu a použití v jiných systémech. Historie prvních nástrojů pro reverzní inženýrství ("Open DWG") a boj o nadvládu nad formáty dodavatelů CAD byl popsán v kapitole

"Strukturovaná data: základ digitální transformace".

Nástroje pro reverzní inženýrství umožňují legitimní získávání dat z uzavřených proprietárních formátů a rozkládají informace ze smíšeného formátu CAD (BIM) na datové typy a formáty požadované uživatelem, což usnadňuje jejich zpracování a analýzu.

Pomocí reverzního inženýrství a přímého přístupu k informacím z databází CAD zpřístupňuje informace, umožňuje otevřená data a otevřené nástroje, stejně jako analýzu dat pomocí standardních nástrojů, vytváření sestav, vizualizace a integraci s dalšími digitálními systémy (obr. 4.1-12).



Obr. 4.1-12 Přímý přístup k datům CAD minimalizuje závislost na softwarových platformách a směruje k přístupu zaměřenému na data.

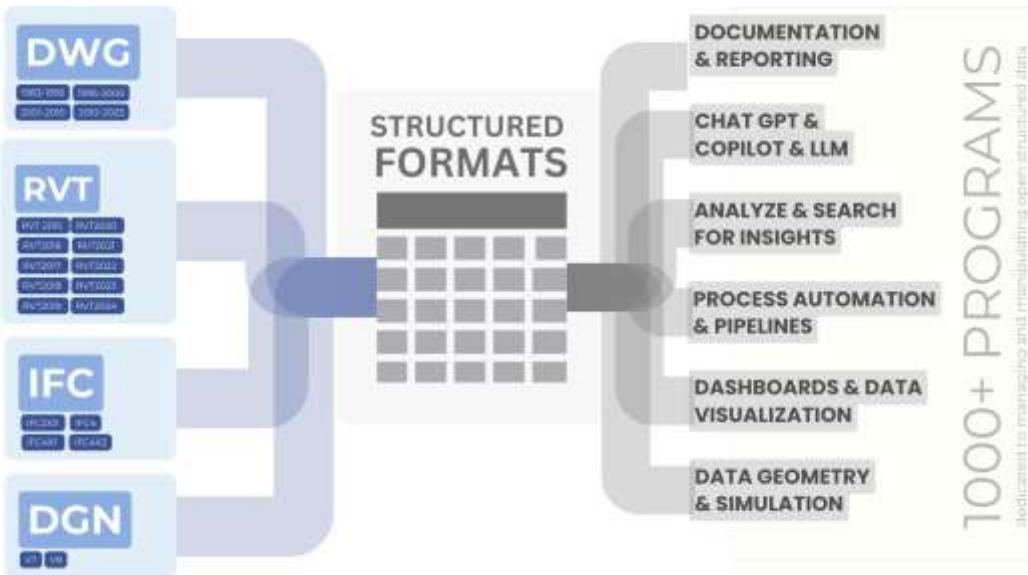
Od roku 1996 pro formát DWG, od roku 2008 pro formát DGN a od roku 2018 pro formát RVT je možné pomocí nástrojů reverzního inženýrství pohodlně a efektivně převádět původně uzavřené formáty dat CAD do libovolných jiných formátů, včetně formátů strukturovaných (obr. 4.1-13). V současné době téměř všechny významné CAD (BIM) a velké inženýrské společnosti na světě používají SDK - nástroje reverzního inženýrství pro extrakci dat z uzavřených formátů CAD (BIM) dodavatele [92].



Obr. 4.1-13 Použití nástrojů reverzního inženýrství umožňuje převést databáze programů CAD (BIM) do libovolného vhodného datového modelu.

Převod dat z uzavřených proprietárních formátů do otevřených formátů a rozdělení smíšených formátů CAD (BIM) na geometrická a metainformační atributová data zjednodušuje proces práce s nimi a zpřístupňuje je pro analýzu, manipulaci a integraci s dalšími systémy (obr. 4.1-14).

Při dnešní práci s daty CAD (BIM) jsme dospěli k tomu, že pro přístup k informacím z formátů CAD nemusíme žádat o povolení dodavatele CAD (BIM).



Obr. 4.1-14 Moderní nástroje SDK umožňují legální převod dat z proprietárních formátů databází CAD (BIM).

Současné trendy v oblasti zpracování dat CAD návrhů jsou i nadále určovány klíčovými hráči na trhu - dodavateli CAD - kteří se snaží posílit svou pozici ve světě dat a vytvářejí nové formáty a koncepce.

Dodavatelé řešení CAD směřují ke strukturovaným datům

Od roku 2024 prochází odvětví projektování a stavebnictví významnou technologickou změnou v oblasti využívání a zpracování dat. Namísto volného přístupu k návrhovým datům se dodavatelé systémů CAD -systémů zaměřují na podporu dalších nových konceptů. Přístupy jako BIM (vytvořený v roce 2002) a otevřený BIM (vytvořený v roce 2012) postupně ustupují moderním technologickým řešením, která začínají prosazovat dodavatelé CAD [93]:

- Přechod na používání "granulovaných" dat, která umožňují efektivní správu informací a přechod k datové analytice.
- Vznik formátu USD a zavedení přístupu Entity-component-system (ECS) pro flexibilní organizaci dat.
- Aktivní využívání umělé inteligence při zpracování dat, automatizaci procesů a analýze dat.
- Vývoj interoperability - lepší interakce mezi různými programy, systémy a databázemi.

Každý z těchto aspektů bude podrobněji rozebrán v šesté části knihy "CAD a BIM: marketing, realita a budoucnost projektových dat ve stavebnictví". V rámci této kapitoly pouze stručně nastíníme obecný vektor změn: hlavní dodavatelé CAD se nyní snaží přehodnotit způsob strukturování informací o projektování. Jedním z klíčových posunů je opuštění klasického souborového modelu ukládání dat ve prospěch granulární, analyticky orientované datové architektury, která poskytuje nepřetržitý přístup k jednotlivým komponentám modelu [93].

Podstatou dění je, že průmysl postupně opouští těžkopádné, specializované a parametrické formáty, které vyžadují geometrická jádra, ve prospěch univerzálnějších, strojově čitelných a flexibilnějších řešení.

Jedním z takových hnacích motorů změn je formát USD (Universal Scene Description), který byl původně vyvinut v odvětví počítačové grafiky, ale díky vývoji platformy NVIDIA Omniverse (a Isaac Sim) pro simulace a vizualizace [93] je již uznáván v technických aplikacích. Na rozdíl od parametrického IFC, USD nabízí jednodušší strukturu a umožňuje popisovat geometrii a vlastnosti objektů ve formátu JSON (obrázek 4.1-15), což usnadňuje zpracování informací a urychluje jejich integraci do digitálních procesů. Nový formát umožňuje ukládat geometrii (kromě BREP -NURBS - podrobněji v 6. části knihy) ve formě polygonů MESH a vlastnosti objektů v JSON, což je výhodnější pro automatizované procesy a práci v cloudových ekosystémech [94].

Někteří dodavatelé CAD a ERP již používají podobné formáty (např. NWD, SVF, CP2, CPIXML), ale většina z nich zůstává uzavřená a nedostupná pro externí použití, což omezuje možnosti integrace a opakovaného použití dat. V této souvislosti může USD hrát stejnou roli jako svého času DXF - otevřenou alternativu k proprietárním formátům, jako je DWG.

General Information				Comparison / Notes
Year of format creation	1991	2016		IFC focuses on construction data, USD on 3D graphics
Creator-developer	TU Munich	Pixar		IFC was founded in Germany, USD in America
Prototypes and predecessors	IGES, STEP	PTEX, DAE, GLTF		IFC evolved from IGES/STEP, USD from PTEX/DAE/GLTF
Initiator in Construction	ADSK	ADSK		ADSK initiated the adoption of both formats in construction
Organizer of the Alliance	ADSK	ADSK		ADSK organized both alliances
Name of the Alliance	bS (IAI)	AOUSD		Different alliances for each format
Year of Alliance Formation	1994	2023		The IFC alliance was formed in 1994, AOUSD for USD in 2023
Promoting in the construction	ADSK and Co	ADSK and Co		ADSK and Co actively promotes both formats in bS (IAI) since the introduction

Purpose and Usage				Comparison / Notes
Purpose	Semantic description and interoperability	Data simplification, visualization unification		IFC for semantics and exchange; USD for simplification and visualization
Goals and Objectives	Interoperability and semantics	Unification for visualization and data processing		IFC focuses on semantics; USD on visualization
Use in Other Industries	Predominantly in construction	In film, games, VR/AR, and now in construction		USD is versatile and used in various fields
Supported Data Types	Geometry, object attributes, metadata	Geometry, shaders, animation, light, and camera		USD supports a wider range of data types suitable for complex visualizations; IFC focuses on construction-specific data

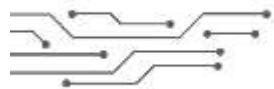
Obr. 4.1-15 USD formát jako snaha dodavatelů CAD splnit požadavek na interoperabilitu a nezávislost návrhových dat na geometrických jádrech.

Přechod hlavních vývojářů na otevřené a zjednodušené formáty USD, GLTF, OBJ, XML (uzavřené NWD, CP2, SVF, SVF2, CPIXML) a podobné formáty (obr. 3.1-17) odráží celosvětový trend a poptávku průmyslu po zjednodušení dat a zvýšení jejich dostupnosti. V příštích letech lze očekávat postupný odklon od složitých parametrických standardů a formátů se závislostí na geometrických jádrech ve prospěch lehčích a strukturovanějších řešení. Tento přechod urychlí digitalizaci stavebnictví, usnadní automatizaci procesů a zjednoduší výměnu dat.

Navzdory strategickým plánům dodavatelů CAD prosazovat nové otevřené formáty mohou odborníci ve stavebnictví získat plný přístup k datům z uzavřených systémů CAD i bez použití nástrojů CAD (BIM), a to pomocí nástrojů reverzního inženýrství.

Všechny tyto trendy nevyhnutelně vedou k přechodu od objemných, monolitických 3D modelů k univerzálním, strukturovaným datům a k používání formátů, které se již dříve osvědčily v jiných odvětvích. Jakmile projektové týmy začnou vnímat modely CAD nejen jako vizuální objekty nebo soubor souborů, ale jako databáze obsahující znalosti a informace, přístup k projektování a správě se dramaticky změní.

Jakmile se týmy naučí získávat strukturovaná data z dokumentů, textů, výkresů a modelů CAD a získají přístup k databázím, je dalším klíčovým krokem modelování dat a zajištění kvality. Právě tento krok do značné míry určuje rychlosť zpracování a transformace informací, které budou nakonec použity k rozhodování v konkrétních aplikačních úlohách.



KAPITOLA 4.2.

KLASIFIKACE A INTEGRACE: SPOLEČNÝ JAZYK PRO STAVEBNÍ DATA

Rychlosť rozhodování závisí na kvalitě dat

Dnešní architektura návrhových dat prochází zásadními změnami. Odvětví se odklání od objemných, izolovaných modelů a uzavřených formátů směrem k flexibilnějším, strojově čitelným strukturám zaměřeným na analýzu, integraci a automatizaci procesů. Samotný přechod na nové formáty však nezaručuje efektivitu - v centru pozornosti je nevyhnutelně kvalita samotných dat.

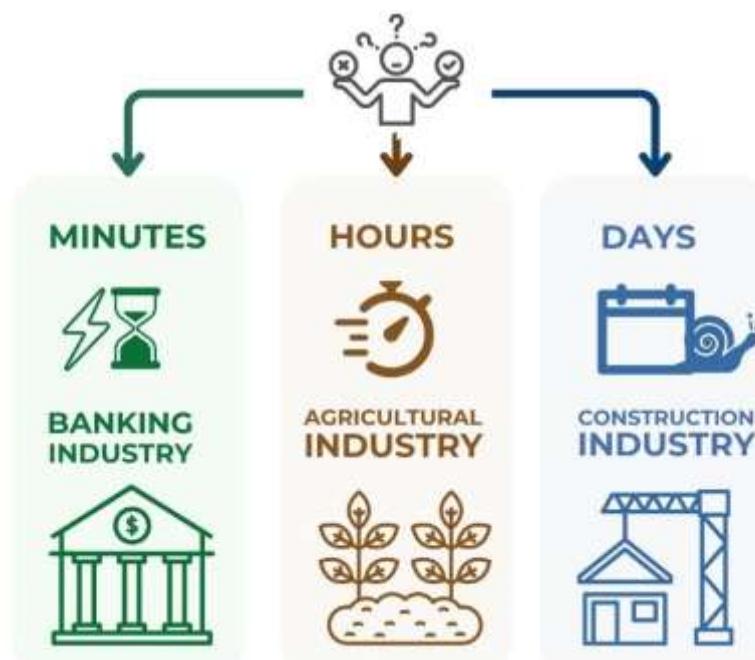
Na stránkách této knihy se hodně mluví o formátech, systémech a procesech. Všechny tyto snahy však nemají smysl bez jednoho klíčového prvku: důvěryhodných dat. Kvalita dat je základním kamenem digitalizace a budeme se k ní vracet i v následujících dílech.

Moderní stavební firmy - zejména ty velké - používají desítky a někdy i tisíce různých systémů a databází (obr. 4.2-1). Tyto systémy musí být nejen pravidelně naplňovány novými informacemi, ale také spolu efektivně komunikovat. Všechna nová data, která vznikají jako výsledek zpracování příchozích informací, jsou do těchto prostředí integrována a slouží k řešení konkrétních obchodních úloh.

A zatímco v minulosti rozhodovali o konkrétních obchodních úkolech vrcholoví manažeři - tzv. HiPPO (obr. 2.1-9) - na základě zkušeností a intuice, dnes, s prudkým nárůstem objemu informací, se tento přístup stává kontroverzním. Nahrazuje jej automatizovaná analytika, která pracuje s daty v reálném čase.

"Tradičně-manuální" diskuse o podnikových procesech na úrovni vedoucích pracovníků se přesunou k provozní analytice, která vyžaduje rychlé reakce na obchodní dotazy.

Doba, kdy účetní, mistři a odhadci ručně vytvářeli zprávy, souhrnné tabulky a přehledy projektových dat v průběhu dnů a týdnů, je minulostí. Dnes se rychlosť a včasnost rozhodování stávají klíčovým faktorem konkurenční výhody.



Obr. 4.2-1 Stavebnictví potřebuje na výpočty a rozhodování dny, na rozdíl od jiných odvětví, kde se tak děje v řádu hodin nebo minut.

Hlavním rozdílem mezi stavebnictvím a digitálně vyspělejšími odvětvími (obrázek 4.2-1) je nízká úroveň kvality a standardizace dat. Zastaralé přístupy k tvorbě, přenosu a zpracování informací zpomalují procesy a vytvářejí chaos. Absence jednotných standardů kvality dat brání zavádění komplexní automatizace

Jedním z hlavních problémů zůstává nízká kvalita vstupních údajů a nedostatek formalizovaných postupů pro jejich přípravu a ověřování. Bez spolehlivých a konzistentních údajů nelze zajistit účinnou integraci mezi systémy. To vede ke zpožděním, chybám a zvýšeným nákladům v každé fázi životního cyklu projektu.

V následujících částech knihy se podrobně venujeme tomu, jak můžete zlepšit kvalitu dat, standardizovat procesy a zkrátit cestu od informací ke kvalitním, ověřeným a konzistentním datům.

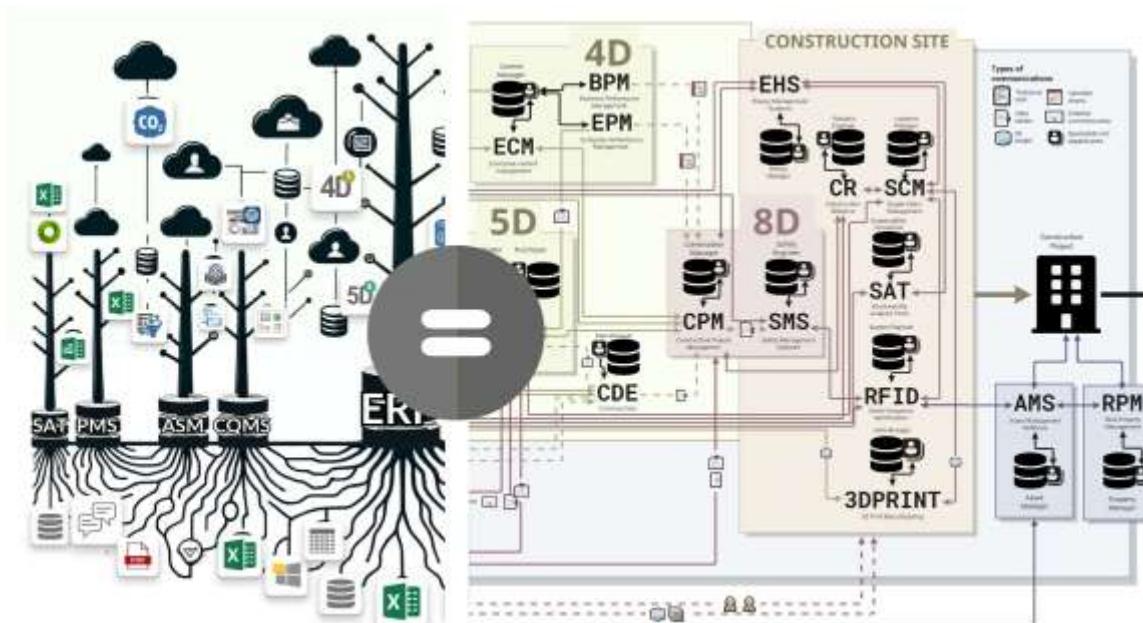
Standardizace a integrace dat

Efektivní správa dat vyžaduje jasnou strategii standardizace. Pouze s jasnými požadavky na strukturu a kvalitu dat lze automatizovat validaci dat, omezit manuální operace a urychlit informované rozhodování ve všech fázích projektu.

V každodenní praxi musí stavební firma denně zpracovávat stovky souborů: e-maily, dokumenty ve formátu PDF, soubory návrhů CAD, data ze senzorů IOT, které je třeba integrovat do podnikových procesů.

Les podnikového ekosystému databází a nástrojů (obrázek 4.2-2) se musí naučit získávat živiny z příchozích multiformátových dat, aby bylo dosaženo výsledků, které společnost chce.

Pro efektivní práci s tokem dat nemusíte nutně najímat armádu manažerů, ale musíte nejprve vypracovat přísné požadavky a standardy pro data a používat vhodné nástroje pro jejich automatickou validaci, harmonizaci a zpracování.



Obr. 4.2-2 Zajištění zdravého podnikového ekosystému vyžaduje kvalitní a včasné zásobování jeho systémů.

Pro automatizaci procesu validace a harmonizace dat (pro následnou automatickou integraci) je třeba začít popisem minimálních požadavků na data pro každý konkrétní systém. Tyto požadavky definují:

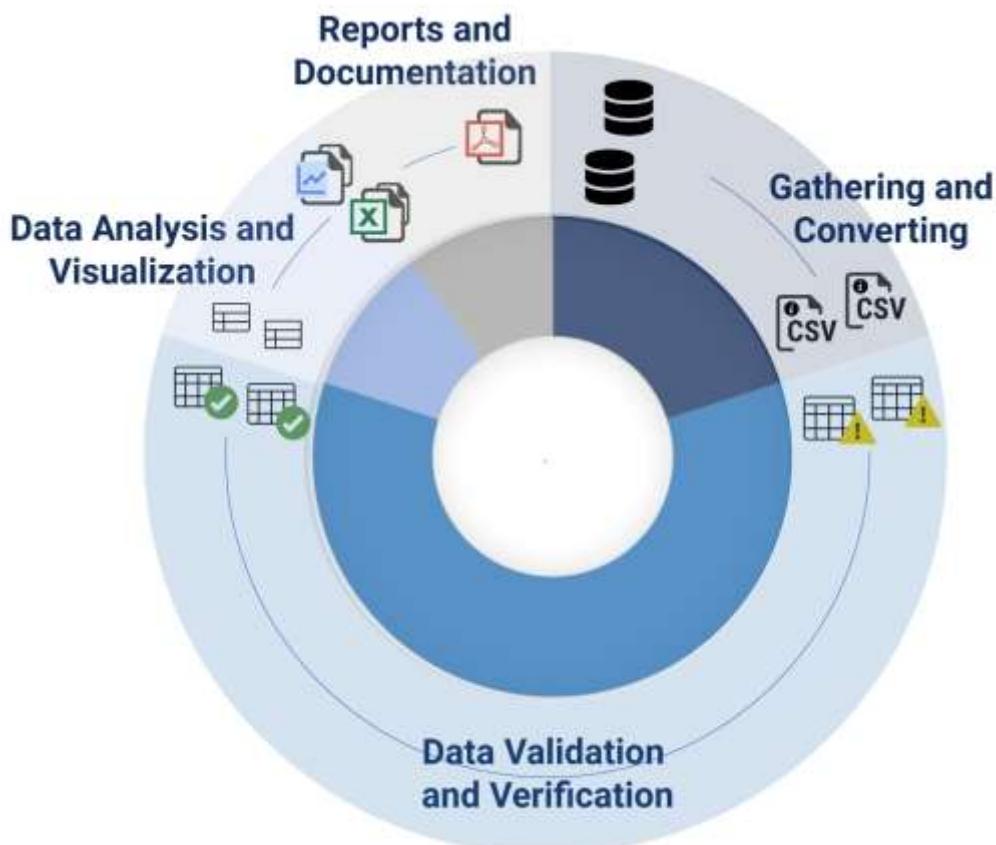
- Co přesně potřebujete získat?
- V jaké formě (struktura, formát)?
- Jaké atributy jsou povinné?
- Jaké tolerance přesnosti a úplnosti jsou přijatelné?

Požadavky na data popisují kritéria kvality, struktury a úplnosti přijímaných a zpracovávaných informací. Například u textů v PDF -dokumentech je důležité zajistit přesné formátování v souladu s oborovými normami (obr. 7.2-14 - obr. 7.2-16). Objekty v CAD -modelu musí mít správné atributy (rozměry, kódy, odkazy na klasifikátory) (obr. 7.3-9, obr. 7.3-10). A pro skenování zakázkového dokumentu je důležité, aby byly všechny relevantní informace včetně klíčových termínů (obr. 4.1-7 - obr. 4.1-10).

Formulace požadavků na data a automatická kontrola jejich dodržování je jedním z časově nejnáročnějších, ale zásadních kroků. Jedná se o časově nejnáročnější krok v podnikových procesech.

Jak bylo uvedeno ve 3. části této knihy, 50 až 90% času profesionálů v oblasti business intelligence (BI) je věnováno spíše přípravě dat než jejich analýze (obrázek 3.2-5). Tento proces zahrnuje sběr, ověřování, validaci, harmonizaci a strukturování dat.

Podle průzkumu z roku 2016 [95] datoví vědci z nejrůznějších širokospektrálních oborů uvedli, že většinu svého pracovního času (přibližně 80%) tráví tím, co je baví nejméně (obrázek 4.2-3): shromažďováním existujících datových souborů a jejich organizováním (sjednocováním, strukturováním). Na kreativní úkoly, jako je hledání vzorců a pravidelností, které povedou k novým poznatkům a objevům, jim tak zbývá méně než 20% času.



Obr. 4.2-3 Ověření a zajištění kvality dat je nejnákladnějším, časově nejnáročnějším a nejsložitějším krokem při přípravě dat pro integraci do jiných systémů.

Úspěšná správa dat ve stavební firmě vyžaduje komplexní přístup, který zahrnuje parametrisaci úloh, formulaci požadavků na kvalitu dat a použití vhodných nástrojů pro jejich automatickou validaci.

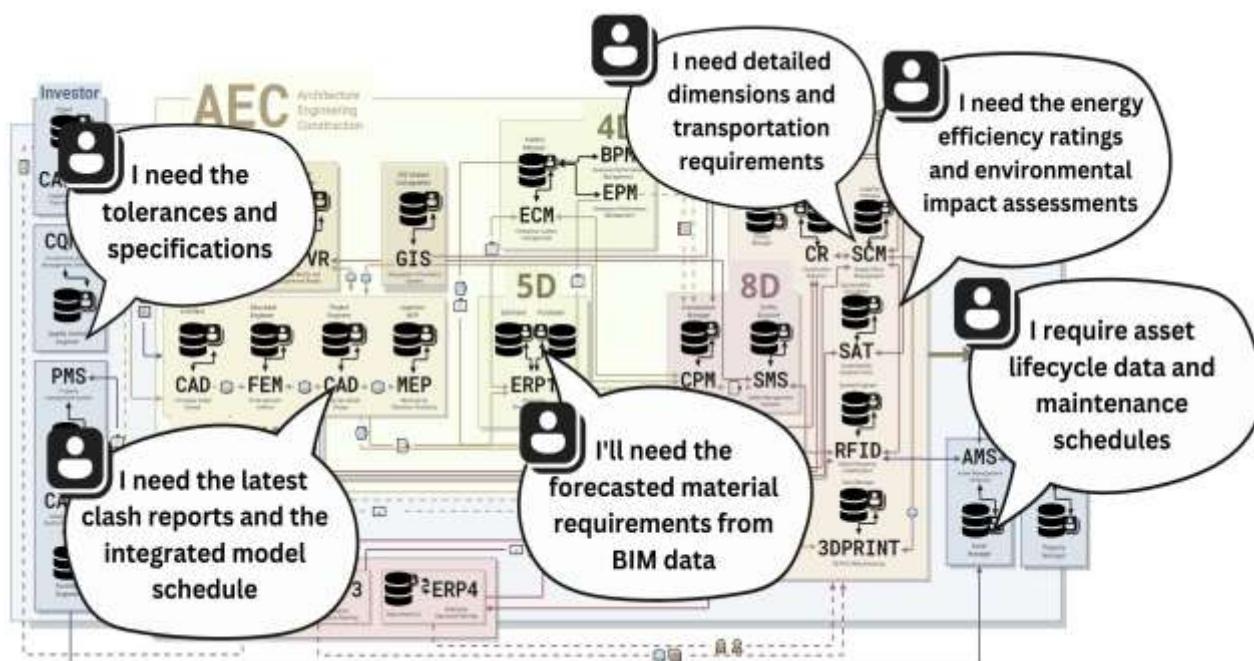
Digitální interoperabilita začíná u požadavků

S rostoucím počtem digitálních systémů ve firmách roste i potřeba konzistence dat mezi nimi. Manažeři odpovědní za různé IT systémy často zjišťují, že nejsou schopni udržet krok s rostoucím objemem informací a různorodostí formátů. Za těchto okolností jsou nuceni požádat specialisty o vytvoření dat ve formě vhodné pro použití v jiných aplikacích a platformách.

To zase vyžaduje, aby se inženýři a pracovníci, kteří se zabývají tvorbou dat, přizpůsobili mnoha požadavkům, často bez transparentnosti a jasného pochopení toho, kde a jak budou data v budoucnu využívána. Nedostatek standardizovaných přístupů k nakládání s informacemi vede k neefektivitě a zvýšeným nákladům ve fázi ověřování, která je často manuální kvůli složitosti a nestandardizované povaze dat.

Otázka standardizace dat není jen otázkou pohodlí nebo automatizace. Jde o přímé finanční ztráty. Podle zprávy společnosti IBM z roku 2016 činí roční ztráta způsobená špatnou kvalitou dat v USA 3,1 bilionu dolarů [96]. Studie MIT a dalších analytických poradenských firem navíc ukazují, že náklady na špatnou kvalitu dat mohou dosáhnout až 15-25% příjmů společnosti [97].

Za těchto podmínek je velmi důležité mít jasně definované požadavky na data a popisy toho, jaké parametry, v jakém formátu a s jakou úrovní podrobnosti by měly být obsaženy ve vytvářených objektech. Bez formalizace těchto požadavků nelze zaručit kvalitu dat a kompatibilitu mezi systémy a fázemi projektu (obr. 4.2-4).

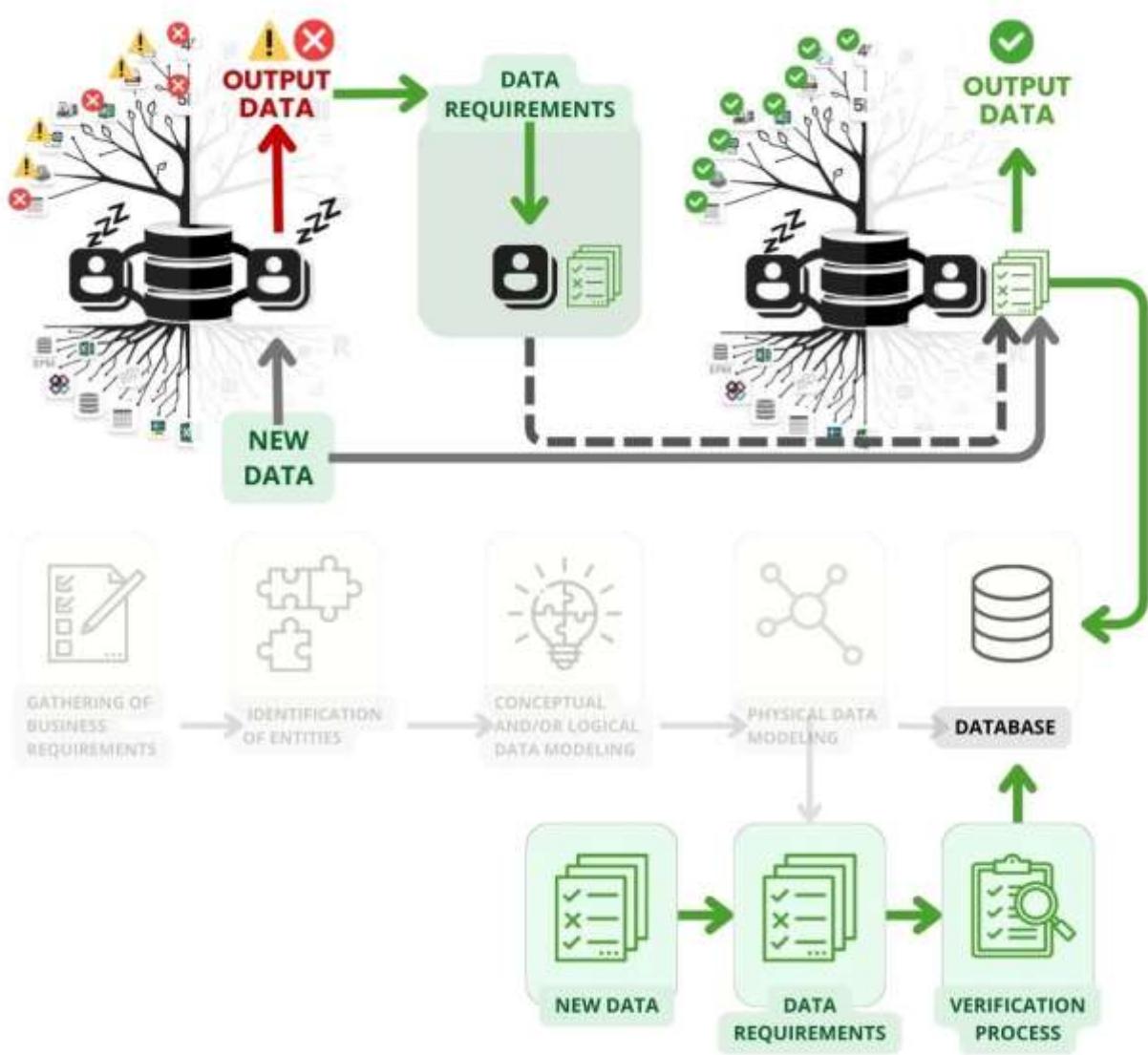


Obr. 4.2-4 Podnikání je založeno na interakci různých rolí, z nichž každá vyžaduje určité parametry a hodnoty, které jsou rozhodující pro naplnění podnikatelských cílů.

Abyste mohli formulovat správné požadavky na data, musíte porozumět obchodním procesům na

úrovni dat. Stavební projekty se liší typem, rozsahem a počtem účastníků a každý systém - ať už jde o modelování (CAD (BIM)), plánování (ERP 4D), kalkulaci nákladů (ERP 5D) nebo logistiku (SCM) - vyžaduje své vlastní jedinečné parametry vstupů (vstupních entit-prvků).

V závislosti na těchto potřebách musí podnikoví manažeři buď navrhnout nové datové struktury, aby splnili požadavky, nebo přizpůsobit stávající tabulky a databáze. Kvalita vytvořených dat bude přímo záviset na tom, jak přesně a správně budou formulovány požadavky (obr. 4.2-5).



Obrázek 4.2-5 Kvalita dat závisí na kvalitě požadavků, které jsou vytvořeny pro konkrétní případy použití dat.

Vzhledem k tomu, že každý systém má své specifické požadavky na data, je prvním krokem při formulování obecných požadavků kategorizace všech prvků zapojených do podnikových procesů. To znamená nutnost rozdělit objekty do tříd a skupin tříd odpovídajících konkrétním systémům nebo

aplikáčním úlohám. Pro každou takovou skupinu se vypracují samostatné požadavky na strukturu, atributy a kvalitu dat.

V praxi se však implementace tohoto přístupu potýká s velkým problémem: chybí společný jazyk pro seskupování dat. Rozdílné klasifikace, duplicitní identifikátory a nekompatibilní formáty vedou k tomu, že si každá společnost, každý software a dokonce i každý projekt vytváří vlastní, izolované datové modely a třídy. Výsledkem je digitální "babylonská věž", kde přenos informací mezi systémy vyžaduje několikanásobné převody na správné datové modely a třídy, často prováděné ručně. Tuto bariéru lze překonat pouze přechodem k univerzálním klasifikátorům a standardizovaným souborům požadavků.

Společný jazyk konstrukce: úloha klasifikátorů v digitální transformaci

V kontextu digitalizace a automatizace kontrolních a zpracovatelských procesů hrají zvláštní roli klasifikační systémy - jakési "digitální slovníky", které zajišťují jednotnost popisu a parametrizace objektů. Klasifikátory tvoří "společný jazyk", který umožňuje seskupovat data podle významu a integrovat data mezi různými systémy, úrovněmi řízení a fázemi životního cyklu projektu.

Nejhmatajelnější dopad třídičů je v ekonomice životního cyklu budovy, kde je nejdůležitějším aspektem optimalizace dlouhodobých provozních nákladů. Studie ukazují, že provozní náklady tvoří až 80% celkových nákladů na vlastnictví budovy, což je třikrát více než počáteční náklady na výstavbu (obr. 4.2-6) [98]. To znamená, že rozhodnutí o budoucích nákladech se z velké části tvoří ve fázi návrhu

Proto by se požadavky od provozních inženýrů (CAFM, AMS, PMS, RPM) měly stát výchozím bodem pro vytváření požadavků na data ve fázi návrhu (obr. 1.2-4). Tyto systémy by neměly být vnímány jako závěrečná fáze projektu, ale jako nedílná součást celého digitálního ekosystému projektu, od konceptu až po demontáž

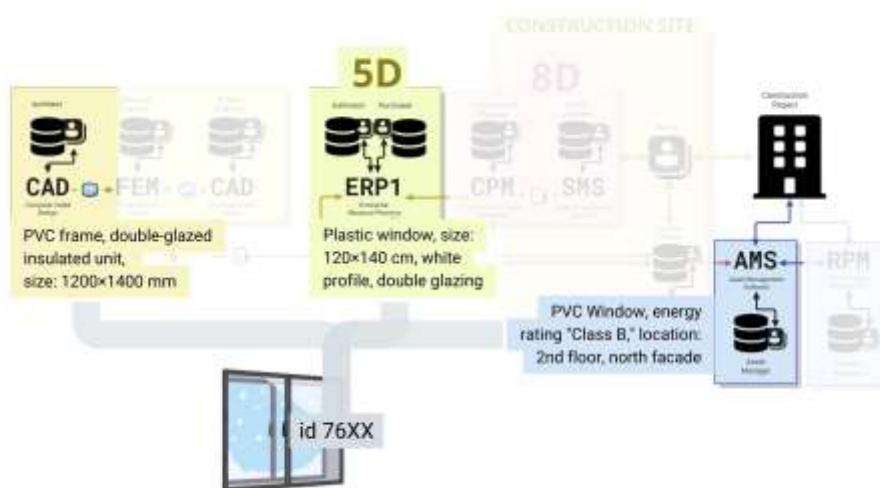
Moderní klasifikátor není jen systém kódů pro seskupování. Je to mechanismus pro vzájemné porozumění mezi architekty, inženýry, odhadci, logistiky, údržby a IT systémů. Stejně jako autopilot automobilu musí jednoznačně a přesně rozpoznávat objekty na silnici, musí digitální stavební systémy a jejich uživatelé prostřednictvím třídy prvků jednoznačně interpretovat stejný prvek projektu pro různé systémy.



Obr. 4.2-6 Náklady na provoz a údržbu trojnásobně převyšují náklady na výstavbu a tvoří 60-80% celkových nákladů životního cyklu budovy (podle [99]).

Úroveň vývoje klasifikátoru přímo souvisí s hloubkou digitalizace společnosti a její digitální vyspělostí. Organizace s nízkou úrovní digitální vyspělosti se potýkají s roztríštěnými daty, nekompatibilními informačními systémy a v důsledku toho i s nekompatibilními a neefektivními klasifikátory. V takových společnostech může mít stejný prvek často v různých systémech různé identifikátory seskupení, což kriticky brání konečné integraci a znemožňuje automatizaci procesů.

Například stejné okno v projektu může být v CAD modelu, systému odhadu a údržby (obr. 4.2-7) označeno různě, a to z důvodu vícerozměrného vnímání prvků různými účastníky procesu. Pro odhadce jsou v kategorii okna důležité prvky, objem a náklady, pro servis údržby dostupnost a udržovatelnost, pro architekta estetické a funkční vlastnosti. V důsledku toho může stejný prvek vyžadovat různé parametry.



Obr. 4.2-7 Při nekonzistentní klasifikaci mezi systémy ztratí prvek při každém kroku přechodu do jiného systému některé informace o atributu.

Vzhledem k obtížnosti jednoznačného definování klasifikace stavebních prvků přiřazují odborníci z

různých oborů často stejnemu prvku neslučitelné třídy. To má za následek ztrátu jednotného pohledu na objekt, což vyžaduje následné manuální zásahy pro harmonizaci různých klasifikačních systémů a vytvoření souladu mezi typy a třídami definovanými různými specialisty.

V důsledku této nejednotnosti často nelze provozní dokumentaci, kterou obdrží oddělení nákupu (ERP) při nákupu stavební položky od výrobce, správně propojit se zatříděním této položky na stavbě (PMIS, SCM). V důsledku toho není pravděpodobné, že by kritické informace byly integrovány do systémů správy infrastruktury a majetku (CAFM, AMS), což způsobuje vážné problémy při uvádění do provozu i při následné údržbě (AMS, RPM) nebo výměně prvku.

Ve společnostech s vysokou digitální vyspělostí hrají klasifikátory roli nervového systému, který spojuje všechny informační toky. Stejný prvek získává jedinečný identifikátor, který umožňuje jeho přenos mezi systémy CAD, ERP, AMS a CAFM -systémy a jejich klasifikátory bez zkreslení nebo ztráty.

Chcete-li vytvořit efektivní klasifikátory, musíte pochopit, jak se data používají. Tentýž inženýr může v různých projektech pojmenovat a klasifikovat prvek různě. Pouze shromažďováním statistik o používání v průběhu let lze vytvořit stabilní klasifikační systém. S tím pomáhá strojové učení: algoritmy analyzují tisíce projektů (obr. 9.1-10) a pomocí strojového učení identifikují pravděpodobné třídy a parametry (obr. 10.1-6). Automatická klasifikace je obzvláště cenná v prostředí, kde není možné provést ruční klasifikaci kvůli objemu dat. Automatické klasifikační systémy budou schopny rozlišit základní kategorie na základě minimálně vyplňených parametrů položek (podrobněji v deváté a desáté části knihy).

Vyvinuté klasifikační systémy se stávají katalyzátorem další digitalizace a vytvářejí základ pro:

- Automatizovaný odhad nákladů a časového harmonogramu projektu.
- Prediktivní analýza potenciálních rizik a konfliktů
- Optimalizace nákupních procesů a logistických řetězců
- Vytváření digitálních dvojčat budov a staveb
- Integrace se systémy chytrých měst a internetu věcí

Čas na transformaci je omezený - s rozvojem technologií strojového učení a počítačového vidění bude v příštích letech vyřešen problém automatické klasifikace, který byl po desetiletí neřešitelný, a stavební a projekční společnosti, které se včas nepřizpůsobí, riskují, že se zopakuje osud vozových parků taxislužby vytlačených digitálními platformami.

Automatizací výpočtů nákladů a harmonogramu, stejně jako velkými daty a strojovým učením se budeme podrobněji zabývat v páté a deváté části knihy. Riziku opakování osudu vozových parků taxislužby a uberizaci stavebnictví se podrobně věnuje desátá část knihy.

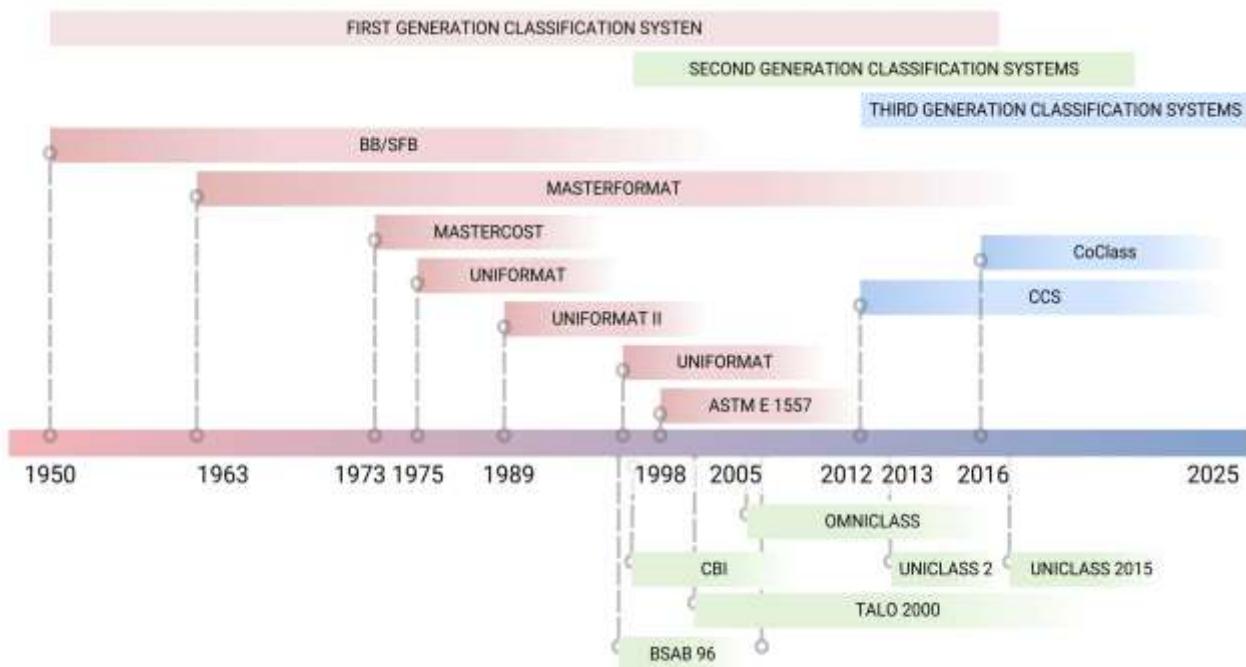
Pro pochopení klíčové role klasifikátorů v digitální transformaci stavebnictví je třeba se obrátit k historii jejich vývoje. Právě historický kontext nám umožňuje uvědomit si, jak se přístupy ke klasifikaci vyvíjely

a jaké trendy určují jejich současný stav.

Masterformat, OmniClass, Uniclass a CoClass: vývoj klasifikačních systémů

Klasifikátory stavebních prvků a prací se historicky vyvíjely ve třech generacích, z nichž každá odrážela úroveň dostupné technologie a aktuální potřeby odvětví v daném časovém období (obrázek 4.2-8):

- **První generace** (od počátku 50. let do konce 80. let) - papírové adresáše, hierarchické klasifikátory používané lokálně (např. Masterformat, SfB).
- **Druhou generaci** (konec 90. let až polovina roku 2010) jsou tabulkové procesory a strukturované databáze implementované v programech Excel a Access (ASTM E 1557, OmniClass, Uniclass 1997).
- **Třetí generace** (od roku 2010 do současnosti) - digitální služby a API - rozhraní, integrace s CAD (BIM), automatizace (Uniclass 2015, CoClass).



Obr. 4.2-8 Tři generace klasifikátorů pro stavebnictví.

V posledních desetiletích došlo ke snížení hierarchické složitosti (obr. 4.2-9) klasifikátorů: zatímco dřívější systémy, jako například OmniClass, používaly až 7 úrovní vnoření pro popis 6887 tříd, moderní řešení, jako například CoClass, se omezují na 3 úrovně se 750 třídami. To usnadňuje práci s daty při zachování potřebné granularity. Uniclass 2015, často používaný jako standard ve Velké Británii, sdružuje 7210 tříd v pouhých 4 úrovních, což je vhodné pro projekty CAD a veřejné zakázky.

Classifier	Table / Objects	Number of classes	Nesting depth
OmniClass	Table 23 Products	6887	7 levels
Uniclass 2015	Pr — Products	7210	4 levels
CoClass, CCS	Components	750	3 levels

Obr. 4.2-9 S každou novou generací klasifikátorů se složitost kategorizace skokově snižuje.

V systémech odhadu stavebních prací v různých zemích může být v důsledku rozdílů v klasifikaci i typický prvek, jako je betonová základová zeď, popsán velmi odlišně (obrázek 4.2-10). Tyto rozdíly odrážejí národní stavební praxi, používané systémy měření, přístupy ke klasifikaci materiálů a regulační a technické požadavky platné v jednotlivých zemích.



Obr. 4.2-10 Stejný prvek se v projektech v různých zemích používá prostřednictvím různých popisů a klasifikací.

Různorodost klasifikací stejných prvků komplikuje mezinárodní spolupráci a činí porovnávání nákladů a rozsahu prací v rámci mezinárodních projektů časově náročným a někdy téměř nemožným. V současné době neexistuje žádný univerzální klasifikátor na celosvětové úrovni - každá země nebo region si vytváří vlastní systémy založené na místních normách, jazyce a obchodní kultuře:

- **CCS** (Dánsko): Systém klasifikace nákladů - systém pro klasifikaci nákladů v průběhu celého životního cyklu zařízení (projekt, výstavba, provoz). Důraz je kladen na logiku provozu a údržby, ale zahrnuje také rozpočet a řízení zdrojů.
- **NS 3451** (Norsko): kategorizuje zařízení podle funkce, konstrukčních prvků a fází životního cyklu. Používá se pro řízení projektů, odhad nákladů a dlouhodobé plánování.
- **MasterFormat** (USA): systém pro strukturování stavebních specifikací do sekcí (např. beton, elektroinstalace, dokončovací práce). Zaměřuje se spíše na obory a typy prací než na funkční prvky (na rozdíl od UniFormatu).
- **Uniclass 2** (Velká Británie): jeden z nejpodrobnějších klasifikátorů, který se používá při zadávání

veřejných zakázek a v projektech BIM. Sjednocuje údaje o objektech, pracích, materiálech a prostorech do jediného systému.

- **OmniClass:** mezinárodní standard (vyvinutý CSI v USA) pro správu informací o objektech od knihoven komponent až po elektronické specifikace. Vhodný pro dlouhodobé ukládání dat, kompatibilní s CAD (BIM) a dalšími digitálními nástroji.
- **COBie:** Construction-Operation Building information exchange je mezinárodní standard pro výměnu dat mezi fázemi návrhu, výstavby a provozu. Zařazen do normy BS 1192-4:2014 jako součást konceptu "BIM -model připravený k použití". Zaměřuje se na přenos informací (např. specifikace zařízení, záruky, kontakty na dodavatele).

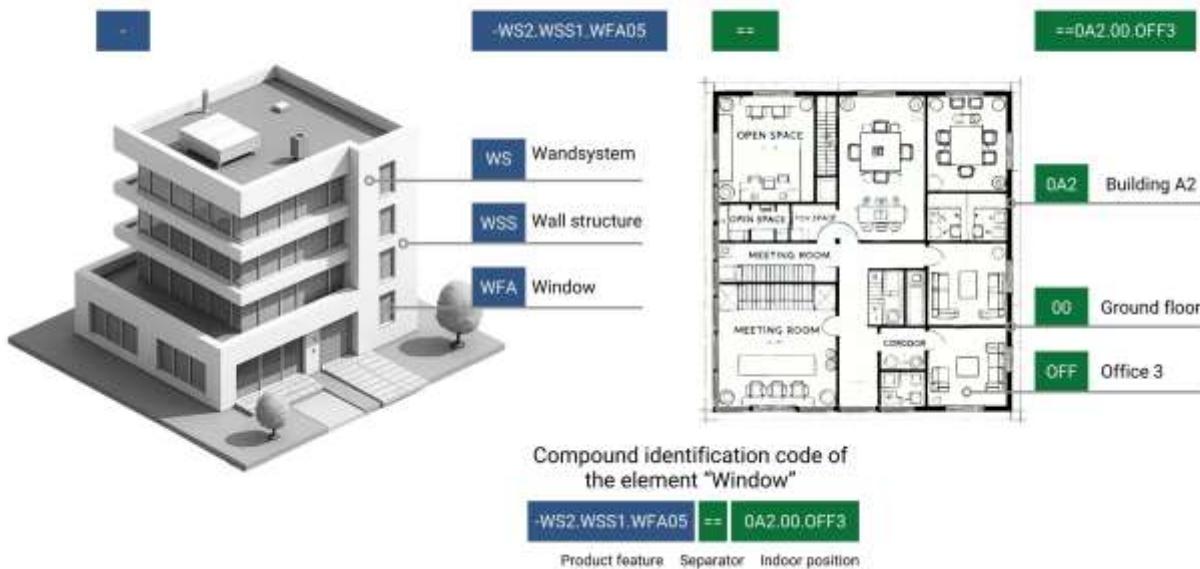
Globalizace stavebnictví pravděpodobně povede k postupnému sjednocování systémů klasifikace stavebních prvků, což výrazně sníží závislost na místních národních normách. Tento proces se může vyvijet po vzoru vývoje internetové komunikace, kde univerzální protokoly pro přenos dat nakonec vytlačily rozdílné místní formáty a zajistily globální interoperabilitu systémů.

Alternativní cestou vývoje může být přímý přechod na automatické klasifikační systémy založené na technologích strojového učení. Tyto technologie, které jsou dnes vyvíjeny především v oblasti autonomní dopravy, mají značný potenciál pro aplikaci na velké soubory dat CAD návrhů (obr. 10.1-6).

Dnes se situace neomezuje pouze na národní seskupení klasifikátorů. Vzhledem k mnoha zvláštnostem, které nejsou na národní úrovni zohledněny, musí každá společnost sjednotit a standardizovat kategorie prvků a zdrojů, s nimiž pracuje.

Tento proces zpravidla začíná v malém měřítku - s lokálními tabulkami objektů nebo interními systémy označování. Strategickým cílem je však přejít ke společnému jazyku pro popis všech prvků, který by byl srozumitelný nejen uvnitř podniku, ale i mimo něj - ideálně harmonizovaný s mezinárodními nebo oborovými klasifikátory (obr. 4.2-8). Tento přístup usnadňuje integraci s externími partnery, digitálními systémy a podporuje tvorbu jednotných end-to-end procesů v rámci životního cyklu objektů.

Před přechodem na automatizaci a škálovatelné IT systémy je nutné buď použít klasifikátory na národní úrovni, nebo vytvořit vlastní, logickou a jednoznačnou strukturu identifikace prvků. Každý objekt - ať už se jedná o okno (obrázek 4.2-11), dveře nebo inženýrský systém - musí být popsán tak, aby byl jednoznačně rozpoznatelný v jakémkoli podnikovém digitálním systému. To je rozhodující při přechodu od plochých výkresů k digitálním modelům, které pokrývají jak fázi návrhu, tak provoz budovy.



Obr. 4.2-11 Příklad složeného identifikátoru okenního prvku budovy na základě klasifikace a polohy v budově.

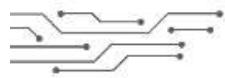
Jedním z příkladů interních klasifikátorů může být vytvoření složeného identifikačního kódu (obrázek 4.2-11). Takový kód kombinuje několik úrovní informací: funkční určení prvku (např. "okno ve stěně"), jeho typ a přesný prostorový odkaz - budova A2, podlaží 0, místnost 3. Taková víceúrovňová struktura umožňuje vytvořit jednotný systém navigace po digitálních modelech a dokumentaci, zejména ve fázích ověřování a transformace dat, kde je vyžadováno jednoznačné seskupení prvků. Jednoznačné rozpoznávání prvků zajišťuje konzistenci mezi jednotlivými odděleními a snižuje riziko duplicit, chyb a ztráty informací.

Dobře vytvořený klasifikátor není jen technický dokument, ale základ digitálního ekosystému společnosti:

- zajišťuje kompatibilitu dat mezi systémy;
- snižuje náklady na vyhledávání a zpracování informací;
- zvyšuje transparentnost a spravovatelnost;
- vytváří základ pro škálování a automatizaci.

Standardizovaný popis objektů pomocí národních klasifikátorů nebo vlastních složených identifikačních kódů se stává základem pro konzistentní data, spolehlivou výměnu informací a následnou implementaci inteligentních služeb - od automatizovaného zadávání veřejných zakázek po digitální dvojčata.

Po dokončení fáze strukturování víceformátových dat a výběru klasifikátoru, který bude použit k rozpoznání a seskupení prvků, je dalším krokem správné modelování dat. Tento proces zahrnuje identifikaci klíčových parametrů, vytvoření logické struktury dat a popis vztahů mezi prvky.



KAPITOLA 4.3.

DATOVÉ MODELOVÁNÍ A CENTRUM EXCELENCE

Modelování dat: konceptuální, logický a fyzický model

Efektivní správa dat (námi dříve strukturovaných a kategorizovaných) není možná bez promyšlené struktury ukládání a zpracování. K zajištění přístupu k informacím a jejich konzistence ve fázi ukládání a zpracování využívají firmy datové modelování, metodiku, která jim umožňuje navrhnout tabulky, databáze a vazby mezi nimi v souladu s obchodními požadavky.

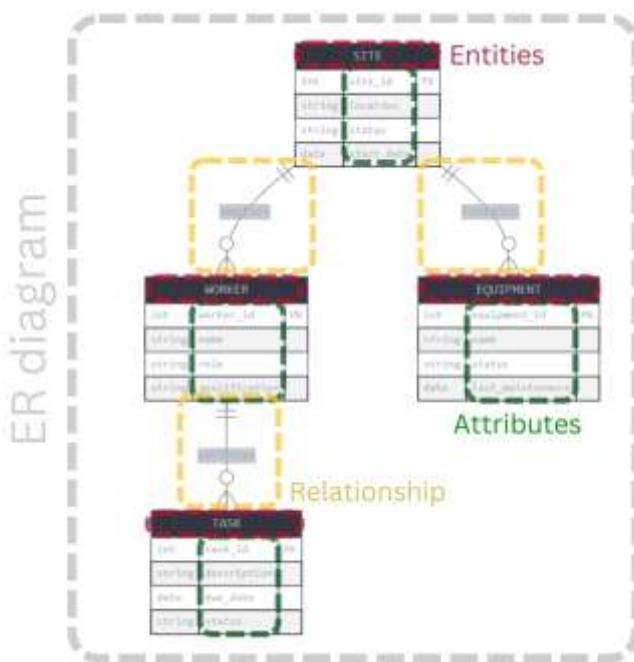
Datové modelování je základem, na kterém je postaven každý digitální ekosystém. Bez popisu systémů, požadavků a datového modelování inženýři a odborníci vytvářející data nevědí a nechápou, kde budou jimi vytvořená data použita.

Stejně jako u stavby budovy, kde nelze začít klást cihly bez plánu, vyžaduje vytvoření systému datového skladu jasnou představu o tom, jaká data se budou používat, jak budou propojena a kdo s nimi bude pracovat. Bez popisu procesů a požadavků ztrácejí inženýři a odborníci vytvářející data přehled o tom, kde a jak budou data v budoucnu využívána.

Datový model slouží jako most mezi obchodem a IT. Umožňuje formalizovat požadavky, strukturovat informace a usnadnit komunikaci mezi zúčastněnými stranami. V tomto smyslu se datové modelování podobá práci architekta, který podle záměru zákazníka vypracuje plán stavby a poté jej předá stavbařům - správcům a vývojářům databáze - k realizaci (vytvoření databáze).

Každá stavební firma tak musí kromě strukturování a kategorizace prvků a zdrojů (obr. 4.2-11) ovládat umění "stavět" databáze (tabulky) a naučit se vytvářet vazby mezi nimi, jakoby spojovat cihly do spolehlivé a pevné zdi znalostí z firemních dat. Klíčové pojmy v datovém modelování (obrázek 4.3-1) zahrnují:

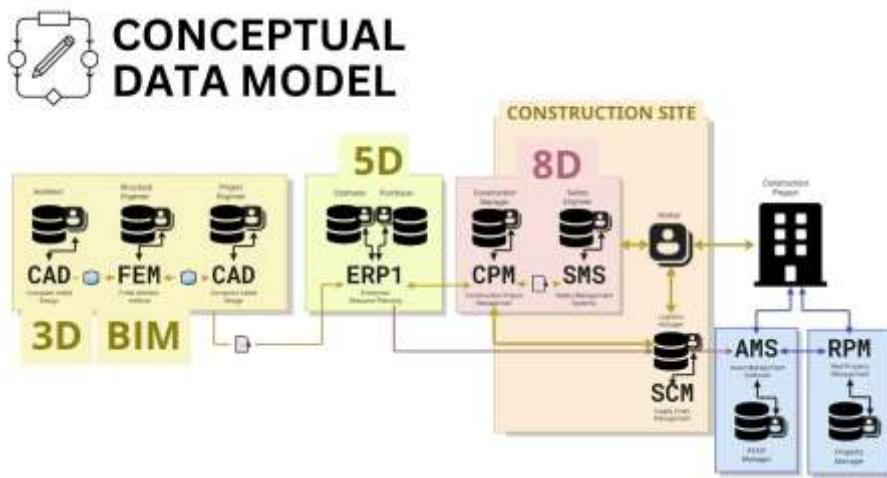
- **Entity** jsou objekty, o nichž je třeba shromažďovat údaje. V počáteční fázi návrhu může být entitou jednotlivý prvek (např. "dveře") a v modelu odhadu to může být skupina prvků zařazených do kategorií (např. "vnitřní dveře").
- **Atributy** jsou charakteristiky entit, které popisují důležité detaily: rozměry, vlastnosti, náklady na montáž, logistiku a další parametry.
- **Vztahy (odkazy)** - ukazují, jak na sebe entity vzájemně působí. Mohou být jednoho z typů: "jeden k jednomu", "mnoho k jednomu", "mnoho k mnoha".
- **ER diagramy (Entity-Relationship diagramy)** jsou vizuální diagramy, které zobrazují entity, atributy a vztahy mezi nimi. ER-diagramy mohou být konceptuální, logické a fyzické - každý z nich odráží jinou úroveň podrobnosti.



Obr. 4.3-1 ER diagram konceptuální struktury databáze s entitami, atributy a vztahy.

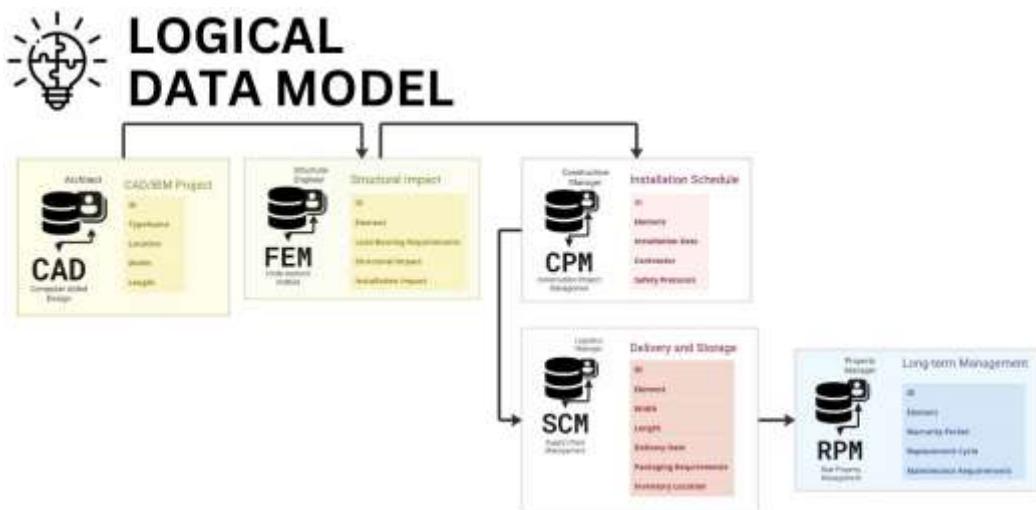
Proces navrhování dat a definování vztahů mezi nimi se tradičně dělí na tři hlavní modely. Každý z nich plní určité funkce a liší se úrovní podrobnosti a stupněm abstrakce při reprezentaci datové struktury:

- **Konceptuální datový model:** tento model popisuje hlavní entity a jejich vztahy, aniž by se zabýval podrobnostmi atributů. Obvykle se používá v počátečních fázích plánování. V této fázi můžeme na základě náčrtů z databází a systémů zobrazit vztahy mezi různými odděleními a specialisty.



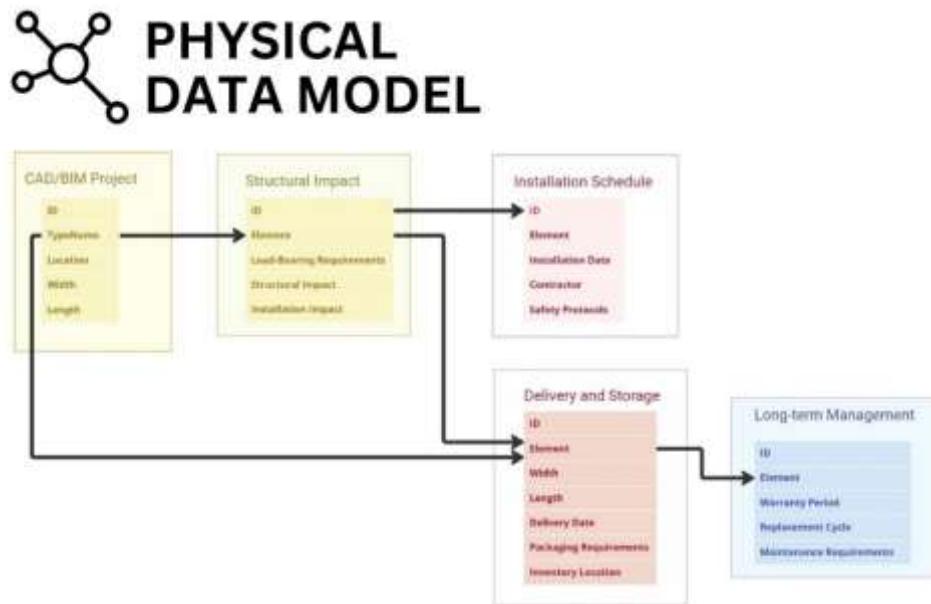
Obr. 4.3-2 Konceptuální diagram popisuje obsah systému: jde o zobrazení vztahů na vysoké úrovni, bez technických detailů.

- **Logický datový model:** Na základě konceptuálního modelu obsahuje logický datový model podrobný popis entit, atributů, klíčů a vztahů, mapování obchodních informací a pravidel.



Obrázek 4.3-3 Logický datový model podrobně popisuje datové typy, vztahy a klíče, ale bez systémové implementace.

- **Fyzický datový model:** Tento model popisuje nezbytné struktury pro implementaci databáze, včetně tabulek, sloupců a vztahů. Zaměřuje se na výkonnost databáze, strategie indexování a fyzické ukládání dat s cílem optimalizovat fyzické nasazení databází.



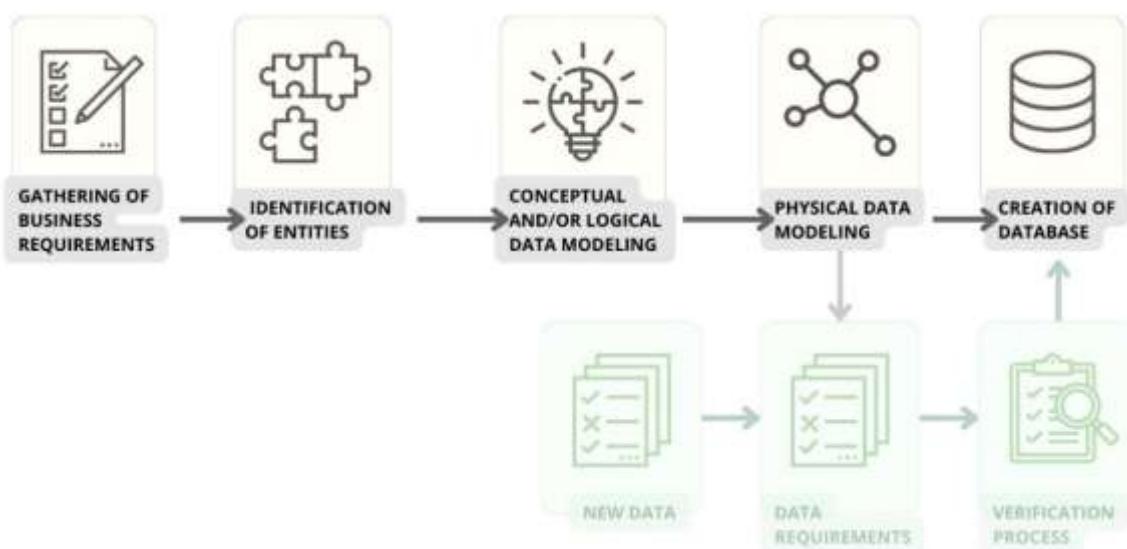
Obrázek 4.3-4 Fyzický datový model definuje, jak bude systém implementován, včetně tabulek a konkrétních detailů databáze.

Při návrhu databází a tabulkových vztahů hraje klíčovou roli při vytváření efektivní architektury systému pochopení úrovní abstrakce.

Efektivní metodika datového modelování umožňuje spojit obchodní cíle s technickou realizací, čímž

se celý procesní řetězec stává transparentnějším a lépe zvládnutelným. Datové modelování není jednorázový úkol, ale proces zahrnující postupné kroky (obrázek 4.3-5):

- **Shromažďování obchodních požadavků:** definují se klíčové úkoly, cíle a informační toky. V této fázi dochází k aktivní interakci s odborníky a uživateli.
- **Identifikace entit:** jsou zdůrazněny hlavní objekty, kategorie a datové typy, které je důležité zohlednit v budoucím systému.
- **Vývoj konceptuálního a logického modelu:** nejprve se zachytí klíčové entity a jejich vztahy, poté atributy, pravidla a podrobná struktura.
- **Fyzické modelování:** je navržena technická implementace modelu: tabulky, pole, vztahy, omezení, indexy.
- **Vytvoření databáze:** posledním krokem je implementace fyzického modelu ve zvoleném DBMS, testování a příprava na provoz.



Obr. 4.3-5 Vytváření databází a systémů pro správu dat pro podnikové procesy začíná tvorbou požadavků a datovým modelováním.

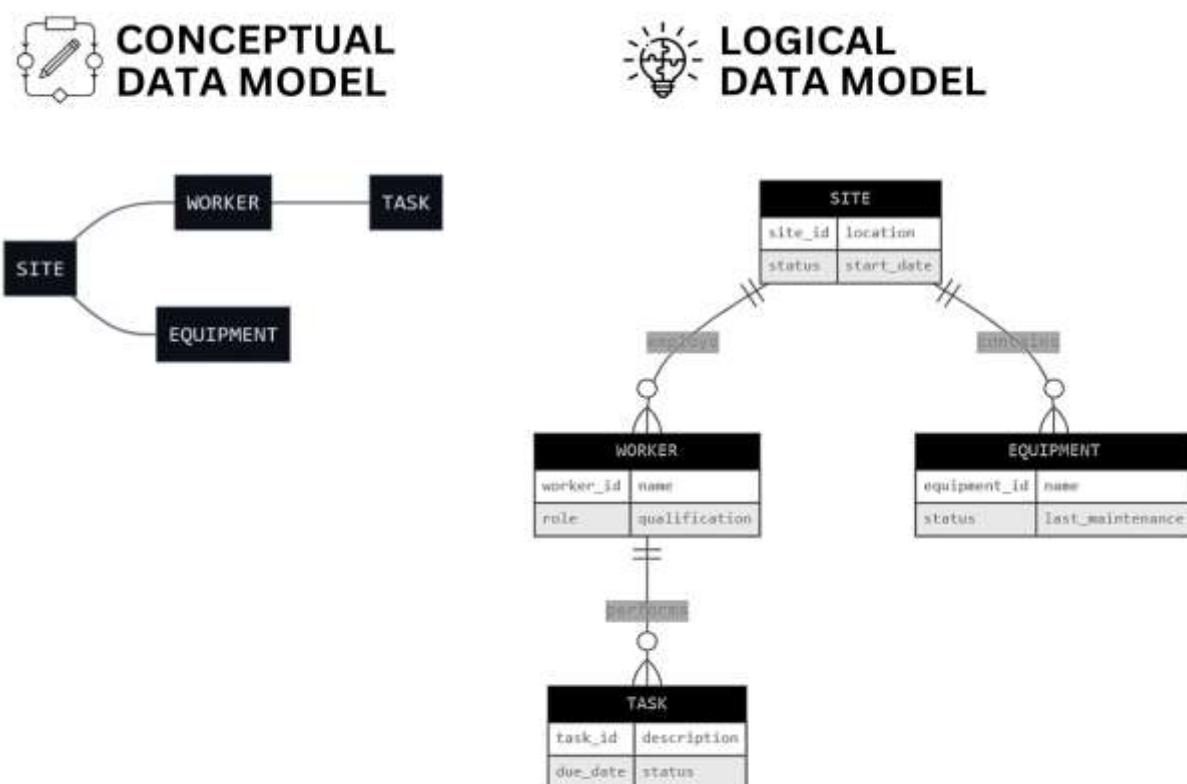
Správně navržené procesy datového modelování umožňují transparentní informační toky, což je důležité zejména v komplexních projektech, jako je řízení stavebních projektů nebo staveniště. Podívejme se, jak může přechod od koncepčního modelu k logickému modelu a následně k fyzickému modelu pomoci zefektivnit procesy.

Praktické modelování dat v kontextu stavebnictví

Vezměme si jako příklad datového modelování úlohu řízení stavby a převedeme požadavky předáka do strukturovaného logického modelu. Na základě základních potřeb řízení staveniště definujeme klíčové

entity pro: staveniště (SITE), pracovníky (WORKER), zařízení (EQUIPMENT), úkoly (TASK) a využití zařízení (EQUIPMENT_USAGE). Každá entita obsahuje sadu atributů, které odrážejí důležité charakteristiky. Například u ÚKOLU to může být popis úkolu, datum splnění, stav, prioritní; u PRACOVNÍKA to může být jméno, jeho role na pracovišti, současné zaměstnání atd.

Logický model vytváří vztahy mezi těmito entitami a ukazuje, jak se vzájemně ovlivňují v reálných pracovních procesech (obrázek 4.3-6). Například vztah mezi pracovištěm a pracovníky naznačuje, že na jednom pracovišti může pracovat mnoho pracovníků, zatímco vztah mezi pracovníky a úkoly odráží, že jeden pracovník může vykonávat více úkolů.

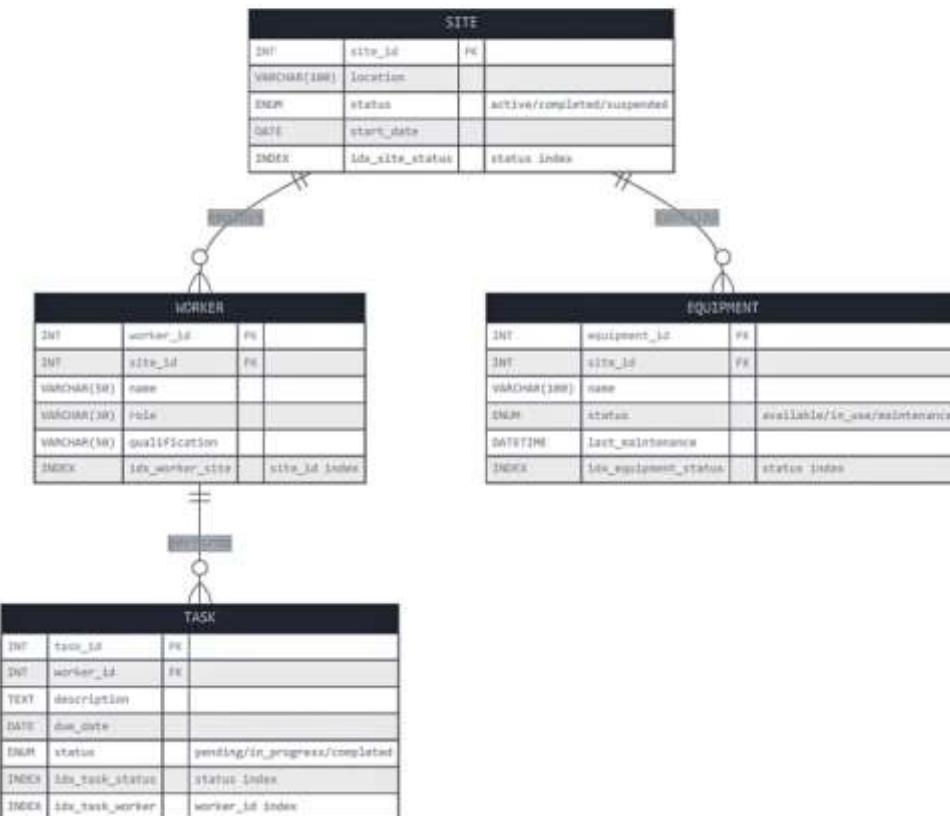


Obr. 4.3-6 Konceptuální a logický datový model vytvořený na základě požadavků předáka pro popis procesů na staveništi.

Při přechodu na fyzický model se přidávají technické implementační detaily: specifické datové typy (VARCHAR, INT, DATE), primární a cizí klíče pro vztahy mezi tabulkami a indexy pro optimalizaci výkonu databáze (obrázek 4.3-7).

Například pro stav by měly být definovány specifické typy s možnými hodnotami a pro zlepšení výkonu vyhledávání by měly být přidány indexy na klíčová pole, jako je status a worker_id. Tím se logický popis systému promění v konkrétní plán implementace databáze, který je připraven k vytvoření a implementaci.

PHYSICAL DATA MODEL



Obrázek 4.3-7 Fyzický datový model popisuje entity staveniště prostřednictvím minimálních požadovaných parametrů.

Fyzický model se často liší od logického modelu. V průměru je rozložení modelovacího času následující: přibližně 50% času se věnuje konceptuálnímu modelu (sběr požadavků, diskuse o procesech, identifikace entit), 10% logickému modelu (specifikace atributů a vztahů) a 40% fyzickému modelu (implementace, testování, přizpůsobení DBMS).

Tato rovnováha je způsobena tím, že konceptuální fáze vytváří základ datové struktury, zatímco logický model pouze specifikuje vztahy a atributy. Fyzický model vyžaduje nejvíce zdrojů, protože právě v této fázi se data implementují do konkrétních platform a nástrojů.

Vytvoření databáze pomocí LLM

Po vytvoření datového modelu a popisu entit pomocí parametrů jsme připraveni vytvořit databáze - úložiště, kam budeme ukládat informace, které přicházejí po fázi strukturování o konkrétních procesech.

Pokusme se vytvořit příklad jednoduché, ale funkční databáze s minimálním množstvím kódu pomocí SQLite na příkladu programovacího jazyka Python. Relační databáze byly podrobně probrány v kapitole "Strukturované relační databáze a dotazovací jazyk SQL".

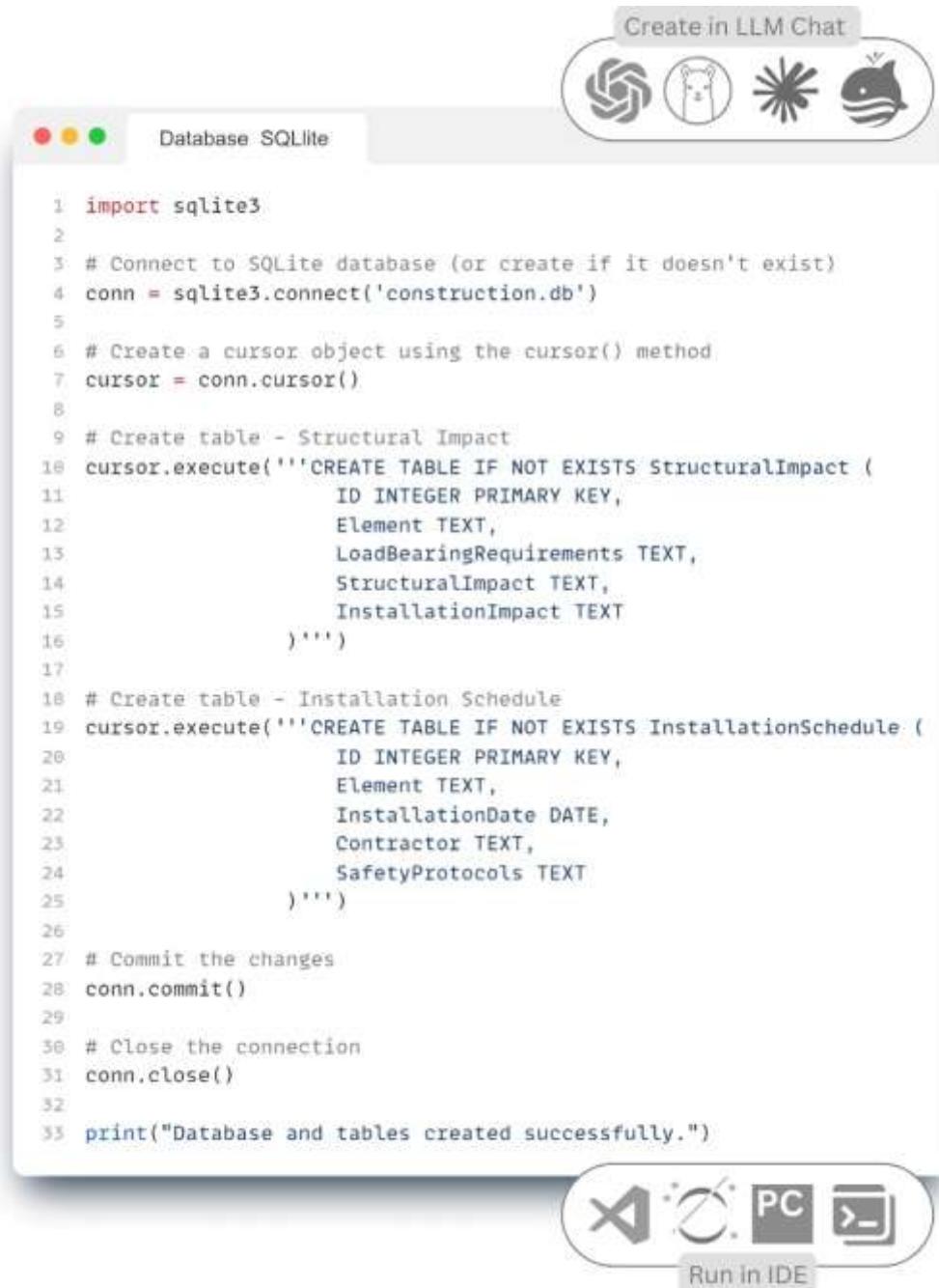
SQLite je lehká, open source, embeddable, multiplatformní databáze, která podporuje základní funkce SQL. SQLite je široce používán v CAD (BIM), ERP, EPM a je přítomen téměř v každém chytrém telefonu.

Požádejme model LLM, aby pro nás vytvořil malou databázi SQLite na základě výše vytvořeného fyzického modelu (obrázek 4.3-4) pro systémy FEM (strukturální dopad) a CPM (montážní plán).

- ❷ Pošlete textovou žádost do chatu LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN nebo do jakéhokoli jiného):

Vytvořte databáze SQLite pro systémy "Konstrukční dopad" a "Plán instalace" podle obrázku v příloze (vložte schéma do chatu jako obrázek (obr. 4.3-4) ↴

■ Odpověď společnosti LLM :



```
1 import sqlite3
2
3 # Connect to SQLite database (or create if it doesn't exist)
4 conn = sqlite3.connect('construction.db')
5
6 # Create a cursor object using the cursor() method
7 cursor = conn.cursor()
8
9 # Create table - Structural Impact
10 cursor.execute('''CREATE TABLE IF NOT EXISTS StructuralImpact (
11                 ID INTEGER PRIMARY KEY,
12                 Element TEXT,
13                 LoadBearingRequirements TEXT,
14                 StructuralImpact TEXT,
15                 InstallationImpact TEXT
16             )''' )
17
18 # Create table - Installation Schedule
19 cursor.execute('''CREATE TABLE IF NOT EXISTS InstallationSchedule (
20                 ID INTEGER PRIMARY KEY,
21                 Element TEXT,
22                 InstallationDate DATE,
23                 Contractor TEXT,
24                 SafetyProtocols TEXT
25             )''' )
26
27 # Commit the changes
28 conn.commit()
29
30 # Close the connection
31 conn.close()
32
33 print("Database and tables created successfully.")
```

Obrázek 4.3-8 Vygenerovaný kód vytvoří z nahraného obrázku databáze a tabulky, do kterých nyní můžeme načíst data.

Tento databázový kód lze spustit v prostředí Pythonu s podporou SQLite, které je obvykle součástí standardního prostředí Pythonu. Lze jej spustit na libovolném lokálním počítači s jazykem Python schopném spouštět skripty a vytvářet soubory. Po spuštění a provedení tento kód vytvoří na pevném disku soubor - databázi SQLite s názvem construction.db (podobně jako soubor aplikace Excel s pracovními listy), který bude obsahovat tabulky "Structural Impact" a "Installation Schedule".

Jakmile je v těchto tabulkách vytvořena nová databáze **construction.db**, lze do těchto tabulek přidávat data pomocí SQL _dotazů nebo importu, což vám později umožní začít vytvářet automatické zpracování dat. Data lze do databáze SQLite importovat ze souborů CSV, z tabulek aplikace Excel nebo je exportovat z jiných databází a úložišť prostřednictvím rozhraní API.

Pro zavedení udržitelného modelování dat a efektivních procesů správy databází potřebuje společnost jasné definovanou strategii a koordinaci mezi technickými a obchodními týmy. V případě různorodých projektů a více zdrojů dat je často obtížné zajistit konzistence, standardizaci a kontrolu kvality na všech úrovních. Jedním z klíčových řešení může být vytvoření specializovaného centra excelence pro datové modelování (COE) v rámci společnosti.

Centrum excelence (CoE) pro datové modelování

Vzhledem k tomu, že data se stávají jedním z klíčových strategických aktiv, musí společnosti dělat víc než jen správně shromažďovat a uchovávat informace - je důležité naučit se data systematicky spravovat. Centrum excelence pro klasifikaci a datové modelování (CoE) je strukturální jednotka, která zajišťuje konzistence, kvalitu a efektivitu veškerého nakládání s daty v organizaci.

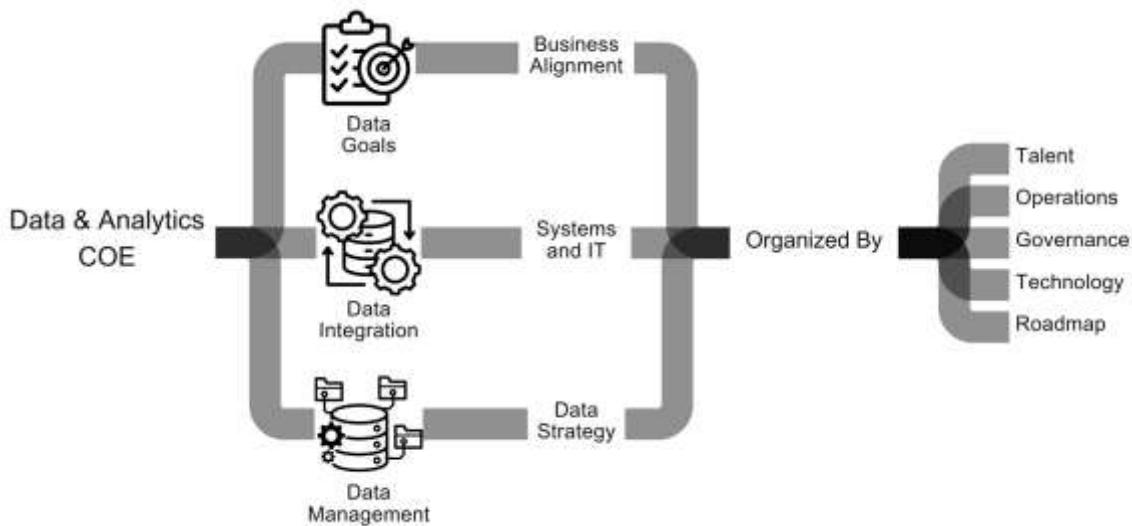
Centrum excelence (CoE) je jádrem odborné podpory a metodickým základem digitální transformace ve firmě. Buduje kulturu založenou na datech a umožňuje organizacím budovat procesy, které se rozhodují na základě strukturovaných, ověřených a reprezentativních dat, nikoli na základě intuice nebo lokálních informací.

Datové centrum excelence je obvykle tvoreno multifunkčními týmy, které pracují podle principu "dvou pizz". Tato zásada, kterou navrhl Jeff Bezos, znamená, že velikost týmu by měla být taková, aby se uživil dvěma pizzami, tj. ne více než 6-10 lidí. Tento přístup pomáhá vyhnout se nadmerné byrokracie a zvyšuje flexibilitu práce. Tým CoE by měl zahrnovat zaměstnance s různými technickými dovednostmi, od datové analýzy a strojového učení až po odborné znalosti v konkrétních oblastech podnikání. Datoví inženýři by díky svým hlubokým technickým znalostem měli nejen optimalizovat procesy a modelovat data, ale také podporovat kolegy tím, že zkrátí čas na rutinní úkoly (obr. 4.3-9).

Stejně jako v přírodě zajišťuje odolnost ekosystémů biodiverzita, v digitálním světě se flexibilita a přizpůsobivost dosahuje díky rozmanitosti přístupů k nakládání s daty. Tato rozmanitost však musí být podložena společnými pravidly a koncepty.

Centrum excelence (CoE) lze přirovnat ke "klimatickým podmínkám" lesního ekosystému, které určují, které typy dat budou prosperovat a které budou automaticky vyřazeny. Vytvořením příznivého "klimatu"

pro kvalitní data usnadňuje CoE přirozený výběr osvědčených postupů a metodik, které se pak stávají standardy organizace.



Obr. 4.3-9 Centrum excelence (CoE) pro data a analytiku sdružuje odborné znalosti klíčových aspektů správy, integrace a strategie dat.

Pro urychlení integračních cyklů a dosažení lepších výsledků by měl CoE poskytnout svým členům dostatečnou míru autonomie při rozhodování. To je důležité zejména v dynamickém prostředí, kde pokusy a omyly, neustálá zpětná vazba a časté vydávání nových verzí mohou přinést významné výhody. Tato autonomie je však účinná pouze tehdy, pokud existuje jasná komunikace a podpora ze strany vyššího vedení. Bez strategické vize a koordinace na nejvyšší úrovni může i ten nejschopnější tým narazit na překážky při realizaci svých iniciativ.

Je to COE nebo vyšší vedení společnosti, které je zodpovědné za to, aby se přístup k datovému modelování neomezil na jeden nebo dva projekty, ale aby byl začleněn do celkového systému řízení informací a obchodních procesů.

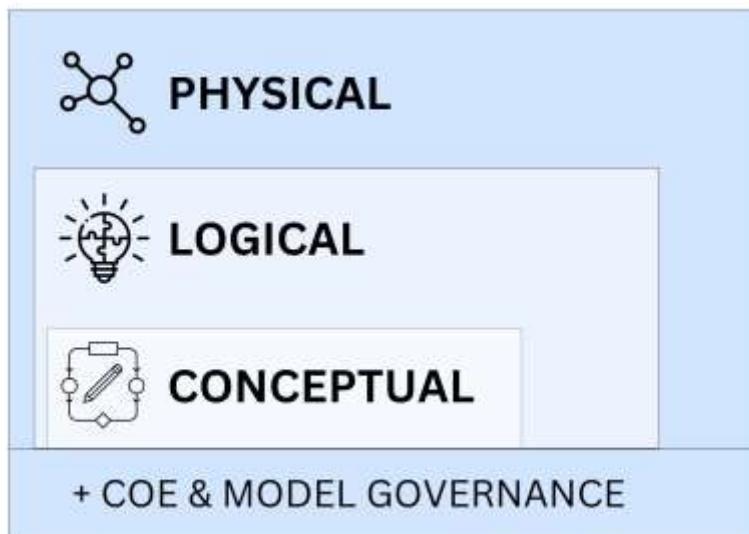
Centrum odbornosti (CoE) je kromě úkolů souvisejících s modelováním a správou dat odpovědné za vývoj společných standardů a přístupů k zavádění a provozu datové infrastruktury. Kromě toho podporuje kulturu neustálého zlepšování, optimalizace procesů a efektivního využívání dat v organizaci (obrázek 4.3-10).

Systematický přístup ke správě dat a modelů v rámci CoE lze zhruba rozdělit do několika klíčových bloků:

- **Standardizace procesů a řízení životního cyklu modelu:** CoE vyvíjí a zavádí metodiky pro sjednocení tvorby a správy datových modelů. To zahrnuje: zavedení strukturálních šablon, metod kontroly kvality a systémů kontroly verzí pro zajištění kontinuity dat ve všech fázích práce.
- **Řízení rolí a přidělování odpovědností:** COE definuje klíčové role v procesu modelování dat. Každému účastníkovi projektu jsou přiděleny jasné definované role a oblasti odpovědnosti,

což usnadňuje týmovou práci a snižuje riziko nekonzistence dat.

- **Kontrola kvality a audit:** efektivní správa stavebních dat vyžaduje průběžné sledování jejich kvality. Zavádějí se automatizované mechanismy pro kontrolu dat, identifikaci chyb a chybějících atributů.
- **Správa metadat a informační architektury:** CoE je odpovědný za vytvoření jednotného systému klasifikace a identifikátorů, standardů pro pojmenování a popis entit, který je rozhodující pro integraci mezi systémy.



Obr. 4.3-10 Modelování dat a řízení kvality dat je jednou z hlavních výzev CoE

Centrum excelence (CoE) pro data není jen skupina odborníků, ale systémový mechanismus, který vytváří novou kulturu založenou na datech a zajišťuje jednotný přístup k práci s daty v celé společnosti. Prostřednictvím kompetentní integrace modelovacích procesů do celkového systému správy informací, standardizace, klasifikace a kontroly kvality dat, pomáhá CoE podnikům neustále zlepšovat jejich produkty a obchodní procesy, rychleji reagovat na změny na trhu a přijímat informovaná rozhodnutí na základě spolehlivých analýz.

Taková centra jsou obzvláště efektivní v kombinaci s moderními principy DataOps - pod krok, který zajišťuje kontinuální dodávání, automatizaci a kontrolu kvality dat. O DataOps budeme více hovořit v části 8, v kapitole "Moderní datové technologie ve stavebnictví".

V následujících kapitolách přejdeme od strategie k praxi - podmíněně se "transformujme" do datového centra: podíváme se na několik příkladů, jak probíhá parametrizace úloh, sběr požadavků a proces automatické validace.



KAPITOLA 4.4.

SYSTEMATIZACE POŽADAVKŮ A OVĚŘOVÁNÍ INFORMACÍ

Shromažďování a analýza požadavků: transformace komunikace do strukturovaných dat.

Shromažďování a správa požadavků je prvním krokem k zajištění kvality dat. Navzdory rozvoji digitálních nástrojů je většina požadavků stále formulována nestrukturovaným způsobem: prostřednictvím dopisů, zápisů z jednání, telefonátů a ústních diskusí. Tato forma komunikace ztěžuje automatizaci, validaci a opakované použití informací. V této kapitole se podíváme na to, jak převést textové požadavky do formálních struktur a zajistit tak transparentnost a systematickost obchodních požadavků.

Výzkum společnosti Gartner "Data Quality: Best Practices for Accurate Insights" zdůrazňuje zásadní význam kvality dat pro úspěšné datové a analytické iniciativy [100]. Upozorňuje, že špatná kvalita dat stojí organizace v průměru nejméně 12,9 milionu dolarů ročně a že spolehlivá a kvalitní data jsou pro vytvoření společnosti založené na datech zásadní.

Absence strukturovaných požadavků vede k tomu, že stejný prvek (entita) a jeho parametry mohou být v různých systémech uloženy v různých variantách. To nejen snižuje efektivitu procesů, ale také vede ke ztrátě času, duplicitě informací a nutnosti opětovného ověřování údajů před jejich použitím. V důsledku toho může i jediné opomenutí - ztracený parametr nebo jediný nesprávně popsaný prvek - zpomalit rozhodování a způsobit neefektivní využívání zdrojů.

Kvůli nedostatku hřebíku se podkova ztratila.

Pro nedostatek podkovy se ztratil kůň.

Pro nedostatek koně se ztratil jezdec.

Pro nedostatek jezdce se ztratila zpráva.

Pro nedostatek poselství byla prohraná bitva.

Pro nedostatek bitvy bylo ztraceno království.

A to vše kvůli nedostatku hřebíku v podkově.

- Přísloví [101]

Analýza a shromažďování požadavků na proces vyplňování a ukládání dat začíná identifikací všech zúčastněných stran. Stejně jako příslovečná ztráta jediného hřebíku vede k řetězci kritických důsledků, tak i v podnikání může ztráta jediné zainteresované strany, přehlédnutý požadavek nebo ztráta byť jediného parametru významně ovlivnit nejen jednotlivý podnikový proces, ale celý ekosystém projektu a organizaci jako celek. Proto je zásadní identifikovat i ty prvky, parametry a role, které se na první pohled zdají nepodstatné, ale později se mohou ukázat jako kritické pro udržitelnost podnikání.

Představme si, že společnost má projekt, v němž klient vznese nový požadavek - "přidat další okno na severní straně budovy". Malého procesu "požadavek klienta na přidání nového okna do stávajícího

"projektu" se účastní architekt, klient, specialista CAD (BIM), stavbyvedoucí, manažer logistiky, ERP - analytik, inženýr kontroly kvality, bezpečnostní inženýr, vedoucí kontroly a manažer nemovitosti.

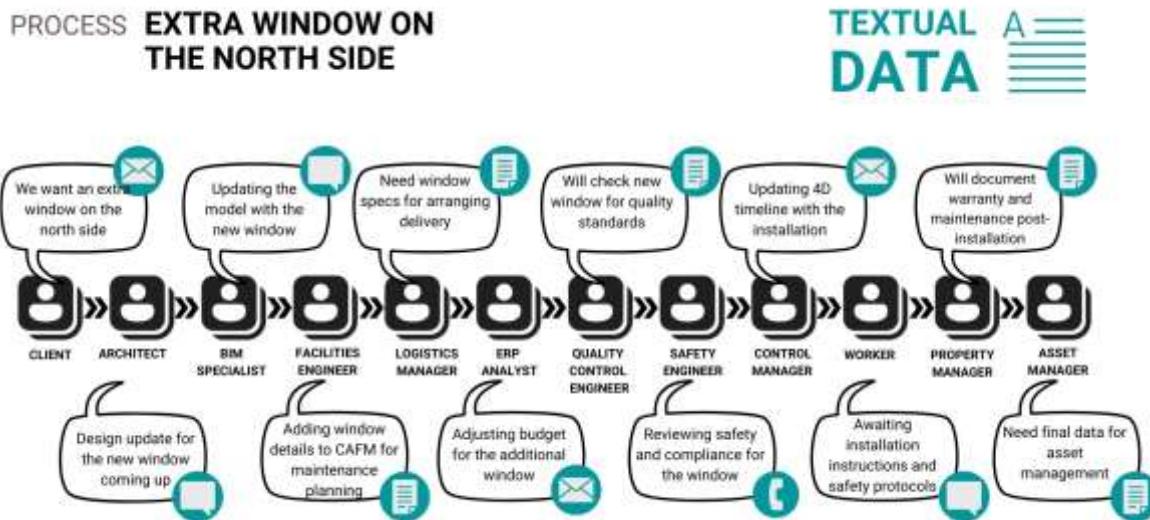
I na malém procesu se mohou podílet desítky různých odborníků. Každý účastník procesu musí rozumět požadavkům specialistů, se kterými je propojen na datové úrovni.

Na úrovni textu (obr. 4.4-1) probíhá komunikace mezi klientem a specialisty v procesním řetězci následujícím způsobem:

- ⦿ **Zákazník:** "Rozhodli jsme se přidat další okno na severní straně pro lepší osvětlení. Lze to realizovat?"
- ⦿ **Architekt:** "Jistě, přepracuji projekt tak, aby obsahoval nové okno, a pošlu aktualizované plány CAD (BIM)".
- ⦿ **Specialista CAD (BIM):** "Obdržel jsem nový projekt. Aktualizuji model CAD (BIM) o další okno a po koordinaci s inženýrem FEM poskytnu přesné umístění a rozměry nového okna".
- ⦿ **Vedoucí stavby:** "Byl přijat nový projekt. Upravujeme termíny instalace 4D a informujeme všechny příslušné subdodavatele".
- ⦿ **Inženýr zařízení (CAFM):** "Zadám údaje 6D na novém okně do systému CAFM pro budoucí správu zařízení a plánování údržby."
- ⦿ **Vedoucí logistiky:** "Potřebuji rozměry a hmotnost nového okna, abych mohl zorganizovat dodávku okna na stavbu."
- ⦿ **ERP -analytik:** "Potřebuji tabulky rozsahu a přesný typ okna pro aktualizaci rozpočtu 5D v našem systému ERP, aby se náklady na nové okno promítly do celkového odhadu projektu."
- ⦿ **Inženýr kontroly kvality:** "Jakmile budou specifikace oken hotové, ujistím se, že splňují naše standardy kvality a materiálu."
- ⦿ **Bezpečnostní inženýr:** "Budu posuzovat bezpečnostní aspekty nového okna, zejména se zaměřením na dodržování předpisů a evakuaci podle systému 8D".
- ⦿ **Manažer kontroly:** "Na základě přesného rozsahu prací z ERP aktualizujeme naši časovou osu 4D tak, aby odrážela instalaci nového okna, a uložíme nová data do systému správy obsahu projektu."
- ⦿ **Pracovník (instalátor):** "Potřebujete pokyny k instalaci, montáži a časovému rozvržení práce. Kromě toho jsou zavedena nějaká zvláštní bezpečnostní pravidla, která musí dodržovat?"
- ⦿ **Správce nemovitosti:** "Po instalaci zdokumentuj informace o záruce a údržbě pro dlouhodobou správu".
- ⦿ **Správce majetku:** "Inženýre zařízení, pošlete prosím konečná data pro sledování majetku a správu životního cyklu."
- ⦿ **Klient:** "Počkejte, možná spěchám a okno nebude potřeba. Možná bych měl udělat balkon."

V takových scénářích, ke kterým dochází často, způsobí i malá změna řetězovou reakci mezi více systémy a rolemi. V takovém případě je téměř veškerá komunikace v počáteční fázi v textové podobě: e-maily, chaty, zápisu z porad (obr. 4.4-1).

V takovém systému textové komunikace pro stavební projekt je velmi důležitý systém právního potvrzení a záznamu všech operací výměny dat a všech učiněných rozhodnutí. To má zajistit, aby každé rozhodnutí, pokyn nebo provedená změna byly právně platné a dohledatelné, což snižuje riziko budoucích "nedorozumění"

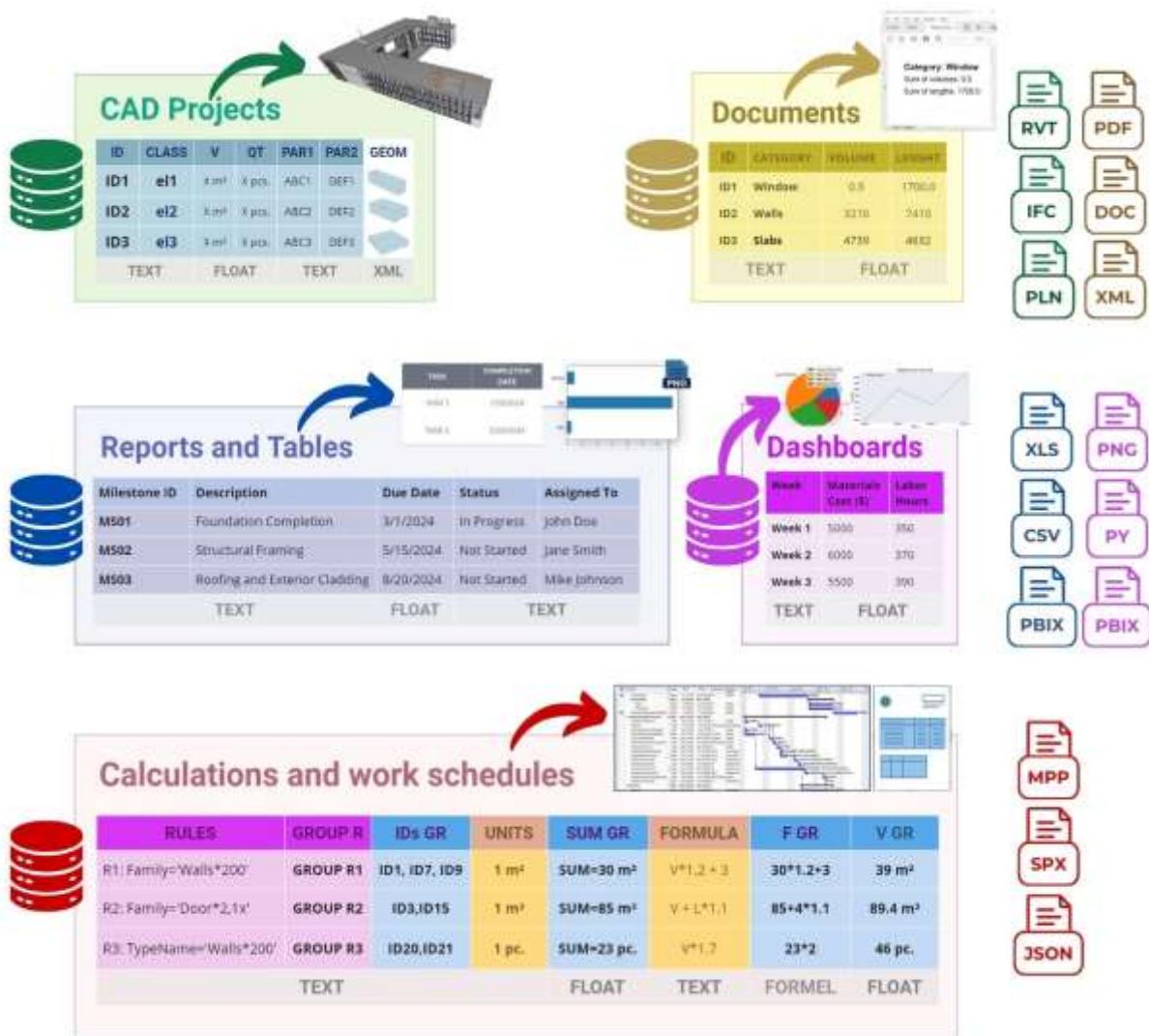


Obr. 4.4-1 Komunikace mezi zadavatelem a dodavatelem v počátečních fázích projektu často obsahuje víceformátová textová data.

Nedostatečná právní kontrola a validace rozhodnutí v příslušných systémech stavebního projektu může vést k vážným problémům pro všechny zúčastněné. Každé rozhodnutí, příkaz nebo změna provedená bez řádné dokumentace a validace může vést ke sporům (a soudním sporům).

Právní konsolidaci všech rozhodnutí v textové komunikaci lze zajistit pouze velkým počtem podepsaných dokumentů, což bude ležet na bedrech vedení, které je povinno zaznamenávat všechny transakce. V důsledku toho, pokud je každý účastník povinen podepisovat dokumenty pro každý úkon, ztrácí systém flexibilitu a stává se byrokratickým bludištěm. Nedostatek potvrzení o transakcích nejenže zpozdí realizaci projektu, ale může vést i k finančním ztrátám a zhoršení vztahů mezi účastníky, a to až k právním problémům.

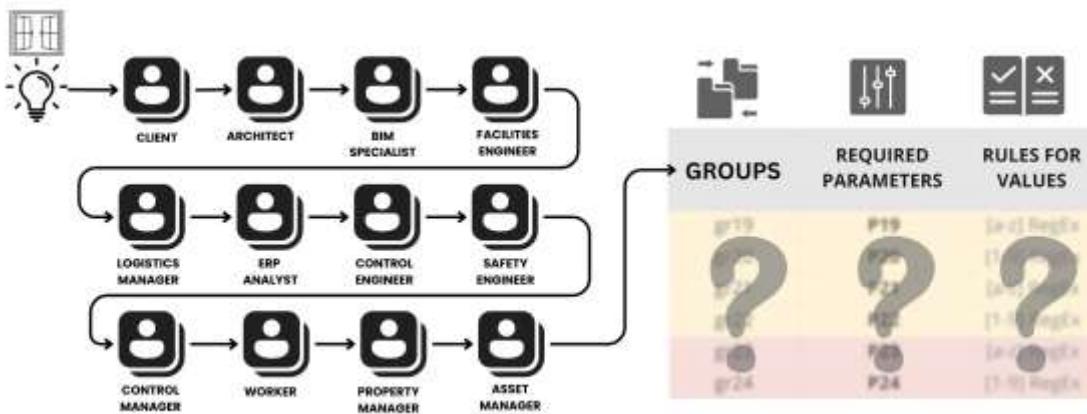
Takový proces schvalování transakcí, který obvykle začíná textovou diskusí, se v následujících fázích postupně vyvíjí ve výměnu dokumentů ve více formátech (obr. 4.4-2), což značně komplikuje komunikaci, která dříve probíhala pouze prostřednictvím textu. Bez jasně definovaných požadavků se automatizace takových procesů, které jsou naplněny víceformátovými daty a velkým množstvím textových požadavků, stává téměř nemožnou.



Obr. 4.4-2 Každý systém v prostředí stavební firmy slouží jako zdroj právně relevantních dokumentů v různých formátech.

Textová komunikace vyžaduje, aby se každý odborník buď seznámil s celou korespondencí, nebo se pravidelně účastnil všech schůzek, aby pochopil aktuální stav projektu.

K překonání tohoto omezení je nutný přechod od textové komunikace ke strukturovanému modelu požadavků. To je možné pouze prostřednictvím systematické analýzy, vizualizace procesů a popisu interakcí ve formě vývojových diagramů a datových modelů (obr. 4.4-3). Stejně jako v případě datového modelování (obr. 4.3-7) jsme se z kontextuálně-ideové úrovni posunuli na konceptuální úroveň přidáním systémů a nástrojů používaných účastníky a vazeb mezi nimi.



Obr. 4.4-3 Abychom se naučili řídit a automatizovat proces validace, je třeba si procesy vizualizovat a požadavky strukturovat.

Prvním krokem při systematizaci požadavků a vztahů je vizualizace všech vazeb a vztahů pomocí koncepčních vývojových diagramů. Konceptuální úroveň nejen usnadní všem účastníkům procesu pochopení celého řetězce procesů, ale také jasně ukáže, proč a pro koho jsou data (a požadavky) v jednotlivých krocích procesu potřeba.

Procesní schémata a účinnost koncepčních rámců

Aby se překlenula propast mezi tradičním a moderním přístupem ke správě dat, musí společnosti vědomě přejít od fragmentovaných textových popisů ke strukturovaným reprezentacím procesů. Vývoj dat - od hliněných tabulek k digitálním ekosystémům - vyžaduje nové nástroje myšlení. A jedním z takových nástrojů je konceptuální modelování pomocí vývojových diagramů. Vytváření vizuálních diagramů - vývojových diagramů, procesních diagramů, diagramů interakcí - umožňuje účastníkům projektu uvědomit si, jak jejich akce a rozhodnutí ovlivňují celý rozhodovací systém.

Pokud procesy vyžadují nejen ukládání dat, ale i jejich analýzu nebo automatizaci, je třeba se začít zabývat tématem tvorby koncepční a vizuální vrstvy požadavků.

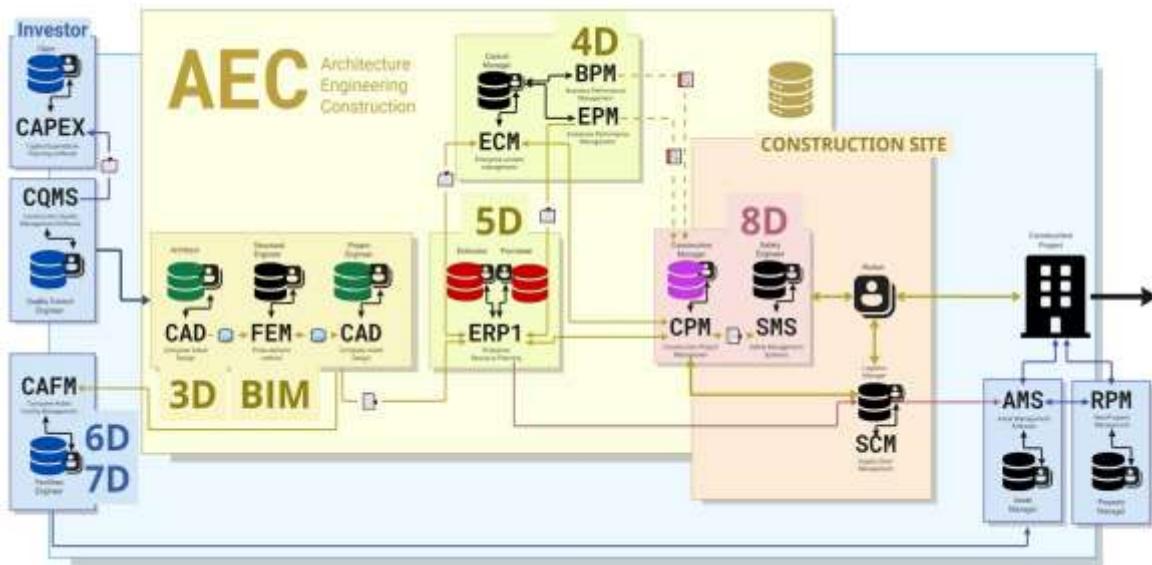
V našem příkladu (obr. 4.4-1) může být každý specialista součástí malého týmu, ale také většího oddělení, zahrnujícího až desítku odborníků pod kontrolou generálního ředitele. Každé oddělení používá specializovanou aplikacní databázi (obr. 1.2-4, např. ERP, CAD, MEP, CDE, ECM, CPM atd.), která je pravidelně aktualizována příchozími informacemi potřebnými k vytváření dokumentů, evidenci právního stavu rozhodnutí a řízení procesů.

Proces transakce se podobá práci starověkých manažerů před 4 000 lety, kdy se k právnímu potvrzení rozhodnutí používaly hliněné tabulky a papyrus. Rozdíl mezi moderními systémy a jejich hliněnými a papírovými předchůdci spočívá v tom, že moderní metody navíc zahrnují proces převodu textových informací do digitální podoby pro další automatické zpracování v jiných systémech a nástrojích.

Vytvoření vizualizace procesu v podobě koncepčních vývojových diagramů pomůže popsát jednotlivé kroky a interakce mezi jednotlivými rolemi, čímž se složitý pracovní postup stane přehledným a jednoduchým.

Vizualizace procesů zajišťuje, že logika procesů je transparentní a přístupná všem členům týmu.

Stejný komunikační proces přidání okna do projektu, který byl popsán formou textu, zpráv (obrázek 4.4-1) a blokového schématu, je podobný koncepčnímu modelu, o kterém jsme hovořili v kapitole o modelování dat (obrázek 4.4-4).



Obr. 4.4-4 Koncepční schéma zobrazuje účastníky projektu jako uživatele databáze, kde jejich dotazy propojují různé systémy.

Přestože jsou koncepční diagramy důležitým krokem, mnoho společností se omezuje na tuto úroveň a domnívá se, že vizuální diagram je pro pochopení procesů dostačující. To vytváří iluzi ovladatelnosti: manažeři na takovém vývojovém diagramu snáze vnímají celkový obraz a vidí souvislosti mezi účastníky a fázemi. Taková schémata však neposkytují jasnou představu o tom, jaká data jsou pro každého účastníka potřebná, v jakém formátu mají být předávána a jaké parametry a atributy jsou pro realizaci automatizace povinné. Konceptuální vývojové schéma se podobá spíše mapě trasy: naznačuje, kdo s kým komunikuje, ale neprozrazuje, co se při těchto interakcích přenáší.

I když je proces podrobně popsán na koncepční úrovni pomocí vývojových diagramů, nezaručuje to jeho účinnost. Vizualizace často zjednoduší práci manažerů a umožňuje jim snadněji sledovat proces pomocí krokových reportů. Pro databázové inženýry však nemusí být konceptuální znázornění dostatečně jasné a nemusí poskytovat jasnou představu o tom, jak proces realizovat na úrovni parametrů a požadavků.

S přechodem na komplexnější datové ekosystémy se počáteční implementace koncepčních a vizuálních nástrojů stává klíčovou pro zajištění toho, aby datové procesy byly nejen efektivní, ale také v souladu se strategickými cíli organizace. Abychom tento proces přidávání oken (obrázek 4.4-1) plně převedli na úroveň požadavků na data, musíme jít o úroveň hlouběji a převést koncepční vizualizaci procesu na logickou a fyzickou úroveň dat, požadovaných atributů a jejich hraničních hodnot.

Strukturované požadavky a regulární výrazy RegEx

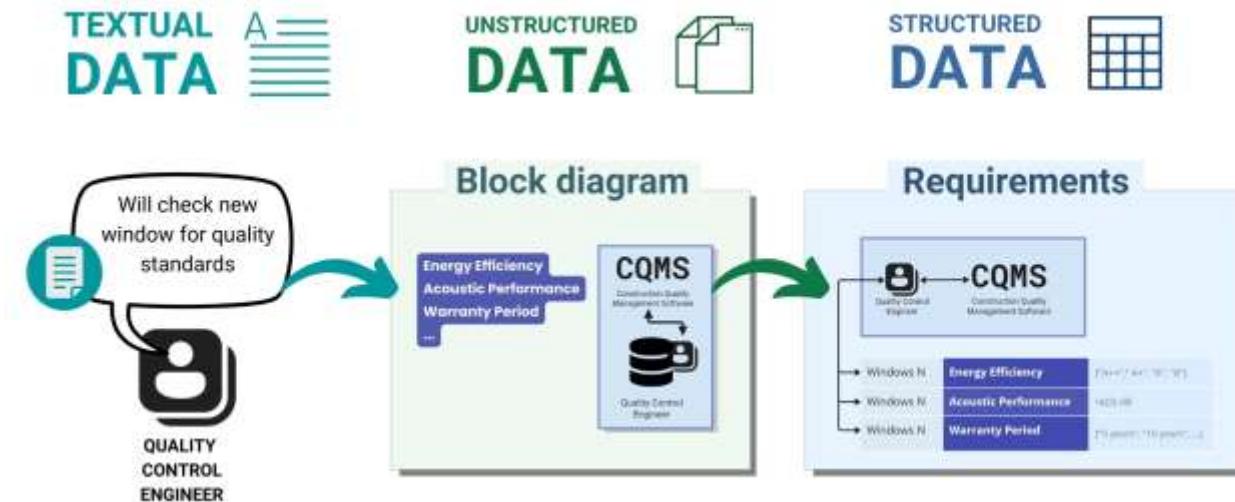
Až 80% dat vytvořených ve firmách je v nestrukturovaných nebo polostrukturovaných formátech [52] - text, dokumenty, dopisy, PDF -soubory, konverzace. Taková data (obr. 4.4-1) je obtížné analyzovat, validovat, přenášet mezi systémy a využívat v automatizaci.

Pro zajištění spravovatelnosti, transparentnosti a automatické validace je nutné převést textové a polostrukturované požadavky do přesně definovaných strukturovaných formátů. Proces strukturování se týká nejen dat (o nichž jsme podrobně hovořili v prvních kapitolách této části knihy), ale také samotných požadavků, které účastníci projektu obvykle formulují ve volné textové podobě v průběhu celého životního cyklu projektu, často bez pomyšlení na to, že tyto procesy lze automatizovat.

Stejně jako jsme již převedli data z nestrukturované textové podoby do podoby strukturované, budeme v pracovním postupu s požadavky převádět textové požadavky do strukturovaného formátu "logické a fyzické vrstvy".

V rámci příkladu přidání okna (obrázek 4.4-1) je dalším krokem popis požadavků na data v tabulkové podobě. Informace pro každý systém používaný účastníky projektu strukturujeme tak, že uvedeme klíčové atributy a jejich hraniční hodnoty

Vezměme si například jeden takový systém (obr. 4.4-5) - systém řízení kvality ve stavebnictví (CQMS), který používá technik kontroly kvality na straně objednatele. S jeho pomocí kontroluje, zda nový prvek projektu - v tomto případě "nové okno" - odpovídá stanoveným normám a požadavkům.



Obr. 4.4-5 Převedení textových požadavků do formátu tabulky s popisem atributů entit zjednodušuje pochopení pro ostatní specialisty.

Jako příklad uveďme některé důležité požadavky na atributy entit typu "okenní systémy" v systému CQMS (obr. 4.4-6): energetická účinnost, akustické vlastnosti a záruční doba. Každá kategorie zahrnuje určité normy a specifikace, které je třeba zohlednit při navrhování a instalaci okenních systémů.



Obr. 4.4-6 Inženýr kontroly kvality by měl zkontrolovat nové prvky typu okna z hlediska energetické účinnosti, zvukové izolace a záručních norem.

Požadavky na údaje, které technik kontroly kvality specifikuje ve formě tabulky, mají například tyto hraniční hodnoty:

- **Třída energetické účinnosti oken** se pohybuje od "A++", která označuje nejvyšší účinnost, po "B", která je považována za minimální přijatelnou úroveň, a tyto třídy jsou reprezentovány seznamem přijatelných hodnot ["A++", "A+", "A", "B"].
- **Akustická izolace oken**, měřená v decibelech a vyjadřující jejich schopnost tlumit hluk z ulice,

je definována pravidelným výrazem \d{2}dB.

- **Atribut "Záruční doba"** pro entitu "Typ okna" začíná na pěti letech, čímž je tato doba nastavena jako minimální povolená při výběru výrobku; zadány jsou také hodnoty záruční doby, jako například **"5 let", "10 let" atd.** nebo logická podmínka **>5 (let)**.

Podle shromážděných požadavků v rámci stanovených atributů neprojdou testem energetické účinnosti nové prvky kategorie nebo třídy oken se známkami nižšími než "B", například "C" nebo "D". Akustická izolace oken v údajích nebo dokumentech pro inženýra kontroly kvality musí být označena dvoumístným číslem následovaným postfixem "dB", například "35 dB" nebo "40 dB", a hodnoty mimo tento formát, například "9 D B" nebo "100 decibelů", nebudou akceptovány (protože neprojdou vzorem pro řetězce RegEx). Záruční doba musí začínat minimálně "5 let" a okna s kratší záruční dobou, jako například "3 roky" nebo "4 roky", nebudou splňovat požadavky, které inženýr kvality popsal ve formátu tabulky.

Pro kontrolu takových hodnot atributových parametrů vůči hraničním hodnotám z požadavků v procesu validace používáme buď seznam povolených hodnot (["A", "B", "C"]), slovníky ({ "A": "H1", "H2": "B": W1, "W2" }), logické operace (např. >, <, <=, >=, ==) pro číselné hodnoty) a regulární výrazy (pro řetězcové a textové hodnoty, jako např. v atributu "Acoustic Performance"). Regulární výrazy jsou mimořádně důležitým nástrojem při práci s řetězcovými hodnotami.

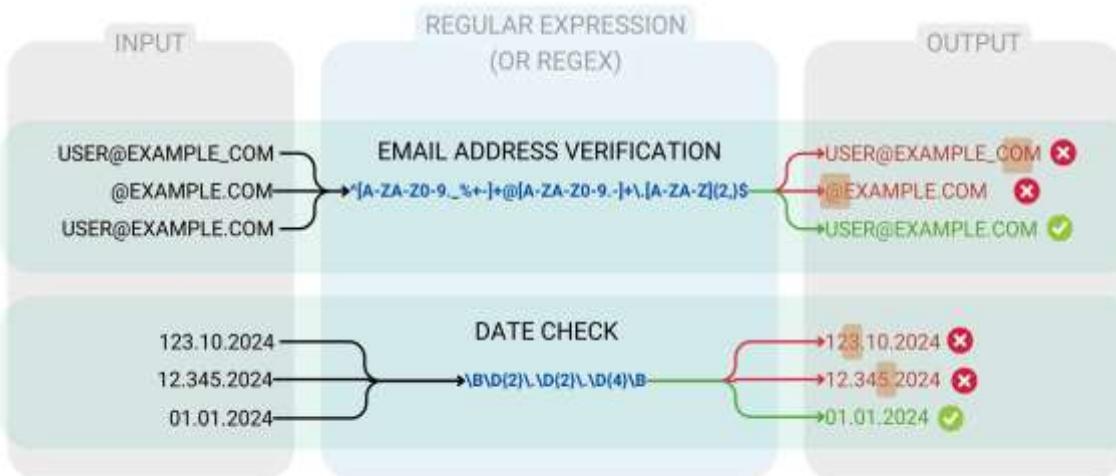
Regulární výrazy (RegEx) se používají v programovacích jazycích, včetně jazyka Python (knihovna Re), k vyhledávání a úpravě řetězců. Regex je jako detektiv ve světě řetězců, který dokáže přesně identifikovat textové vzory v textu.

V regulárních významech se písmena popisují přímo pomocí odpovídajících znaků abecedy, zatímco čísla lze reprezentovat pomocí speciálního znaku \d, který odpovídá libovolné číslici od 0 do 9. Hranaté závorky se používají k označení rozsahu písmen nebo číslic, např. [a-z] pro libovolné malé písmeno latinské abecedy nebo [0-9], které odpovídá \d. Pro nečíselné a nepísmenné znaky se používá \D a \W.

Oblíbené případy použití RegEx (obrázek 4.4-7):

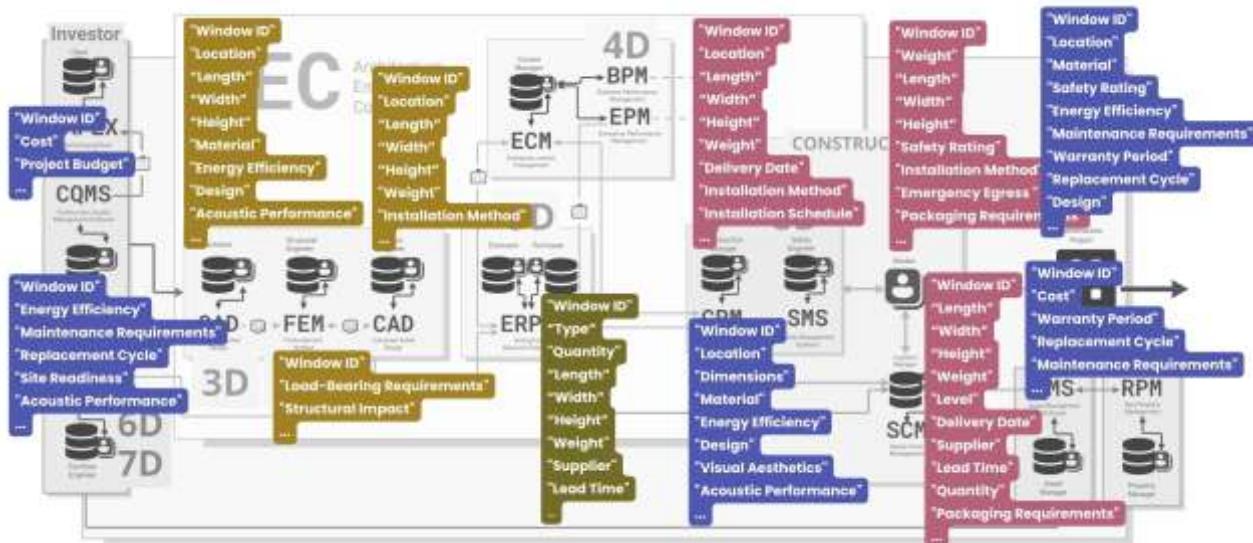
- **Ověření e-mailové adresy** : Chcete-li ověřit, zda je řetězec platnou e-mailovou adresou, můžete použít šablonu "`^ [a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}$`".
- **Extrakce data**: Šablonu "`\b\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{4}\b`" lze použít k extrakci data z textu ve formátu DD.MM.RRRR.
- Ověřování telefonních čísel: pro ověření telefonních čísel ve formátu +49(000)000-0000 bude vzor vypadat takto: "`+\d{2}\(\d{3}\)\)\d{3}-\d{4}`".

Převodem požadavků inženýra kvality do formátu atributů a jejich hraničních hodnot (obr. 4.4-6) jsme je z původní textové podoby (rozhovory, dopisy a regulační dokumenty) převedli do uspořádané a strukturované tabulky, čímž jsme umožnili automatickou kontrolu a analýzu všech příchozích dat (např. nových prvků kategorie Okno). Přítomnost požadavků umožňuje automaticky vyřadit údaje, které nebyly zkontrolovány, a zkontrolované údaje automaticky přenést do systémů k dalšímu zpracování.



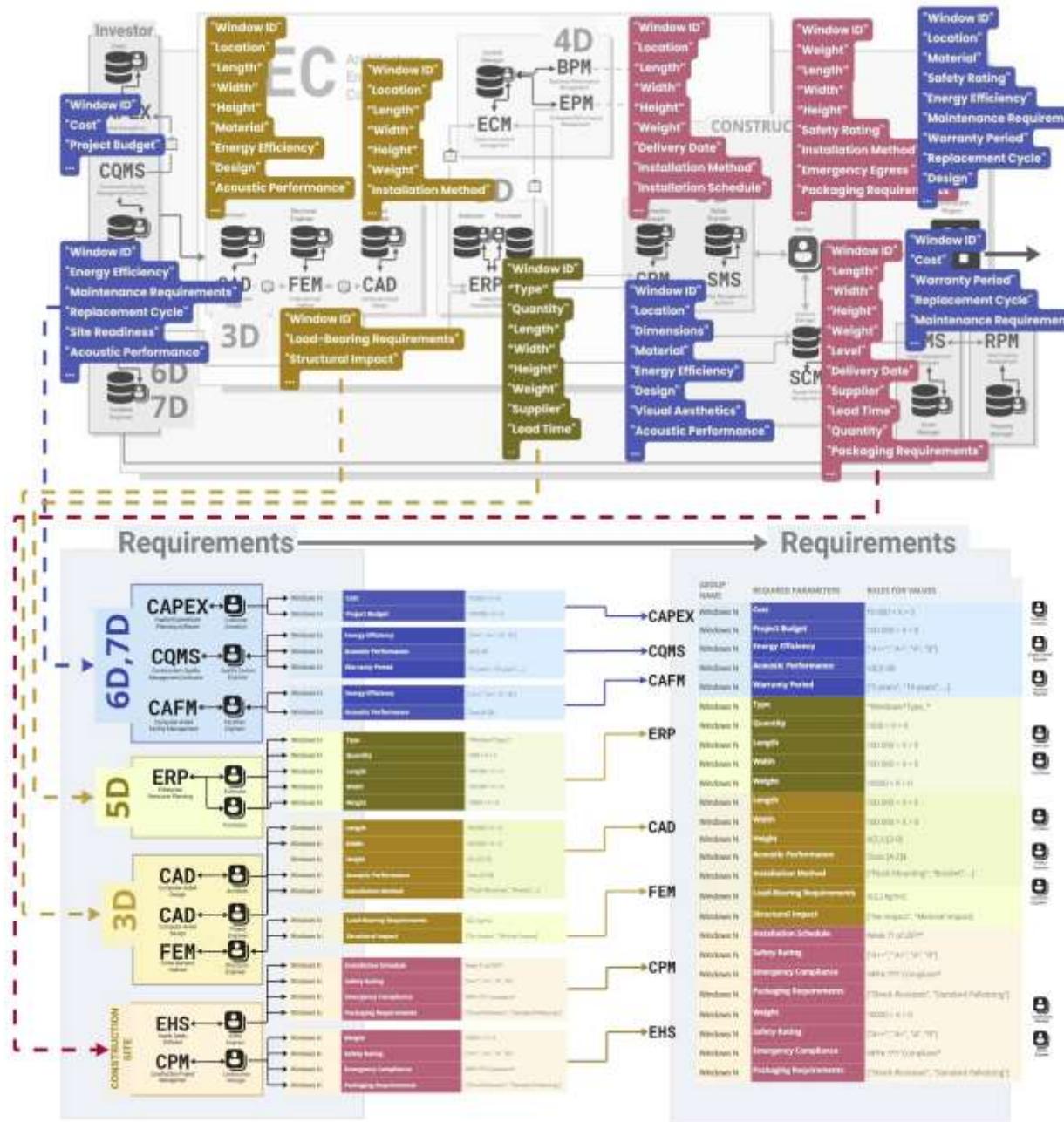
Obrázek 4.4-7 Použití regulárních výrazů je mimořádně důležitým nástrojem v procesu validace textových dat.

Nyní, když přejdeme z konceptuální na logickou úroveň práce s požadavky, převedeme všechny požadavky všech specialistů v našem procesu instalace nového okna (obr. 4.4-4) na uspořádaný seznam ve formátu atributů a tyto seznamy s potřebnými atributy přidáme do našeho vývojového diagramu pro každého specialistu (obr. 4.4-8).



Obrázek 4.4-8 Na úrovni logických procesů jsou atributy, které zpracovávají jednotliví specialisté, přidány do jejich příslušných systémů.

Přidáním všech atributů do jedné společné procesní tabulky transformujeme informace dříve prezentované jako text a dialog na konceptuální úrovni (obr. 4.4-1) do strukturované a systematické podoby tabulek fyzické úrovni (obr. 4.4-9).



Obrázek 4.4-9 Převod nestrukturovaného odborného dialogu do strukturovaných tabulek pomáhá pochopit požadavky na fyzické úrovni.

Nyní je třeba požadavky na data sdělit specialistům, kteří vytvářejí informace pro konkrétní systémy. Pokud například pracujete v databázi CAD, měli byste před zahájením modelování prvků shromáždit všechny potřebné parametry na základě scénářů konečného použití dat. Obvykle se začíná provozní fází, následuje staveniště, oddělení logistiky, oddělení odhadu, oddělení statických výpočtů atd. Teprve poté, co zohledníte požadavky všech těchto oblastí, můžete začít vytvářet data - na základě shromážděných parametrů. To vám v budoucnu umožní automatizovat ověřování a předávání dat v rámci celého řetězce.

Pokud nová data splňují požadavky, jsou automaticky integrována do datového ekosystému společnosti a přechází přímo k uživatelům a systémům, pro které byla určena. Ověřování dat podle atributů a jejich hodnot zajišťuje, že informace splňují požadované standardy kvality a jsou připraveny k použití ve firemních scénářích.

Požadavky na údaje byly definovány a nyní, před zahájením ověřování, je třeba vytvořit, získat nebo shromáždit údaje, které mají být ověřeny, nebo zaznamenat aktuální stav informací v databázích, aby mohly být použity v procesu ověřování.

Sběr dat pro proces ověřování

Před zahájením validace je důležité zajistit, aby byla data k dispozici ve formě vhodné pro proces validace. To neznamená pouze mít informace k dispozici, ale připravit je: data musí být shromážděna a transformována z nestrukturovaných, volně strukturovaných, textových a geometrických formátů do strukturované podoby. Tento proces je podrobně popsán v předchozích kapitolách, kde byly diskutovány metody transformace různých typů dat. Výsledkem všech transformací je, že příchozí data mají podobu otevřených strukturovaných tabulek (obr. 4.1-2, obr. 4.1-9, obr. 4.1-13).

S požadavky a strukturovanými tabulkami s potřebnými parametry a hraničními hodnotami (obr. 4.4-9) můžeme zahájit validaci dat - buď jako jediný automatizovaný proces (Pipeline), nebo jako postupnou validaci každého příchozího dokumentu.

Pro zahájení kontroly je nutné buď přijmout nový soubor jako vstup, nebo opravit aktuální stav dat - vytvořit snímek nebo exportovat aktuální a příchozí data, případně nastavit připojení k externí nebo interní databázi. V uvažovaném příkladu se takový snímek vytvoří automatickým převodem dat CAD z do strukturovaného formátu zaznamenaného například v pátek 29. března 2024 ve 23:00:00, poté co všichni projektanti odešli domů.



Obrázek 4.4-10 Snímek databáze CAD (BIM) zobrazující aktuální informace o atributech nové entity třídy "Okno" v aktuální verzi modelu projektu.

Díky nástrojům pro reverzní inženýrství, o nichž pojednává kapitola "Převod dat CAD (BIM) do strukturované podoby", lze tyto informace z různých nástrojů CAD (BIM) a editorů uspořádat do samostatných tabulek (obr. 4.4-11) nebo je spojit do jedné společné tabulky kombinující různé části

projektu (obr. 9.1-10).

Taková tabulka - databáze zobrazuje jedinečné identifikátory oken a dveří (atribut ID), názvy typů (TypeName), rozměry (Width, Length), materiály (Material), jakož i ukazatele energetické a akustické náročnosti a další charakteristiky. Takovou tabulku vyplněnou v programu CAD (BIM) shromažďuje projektant z různých oddělení a dokumentů a vytváří tak informační model projektu.



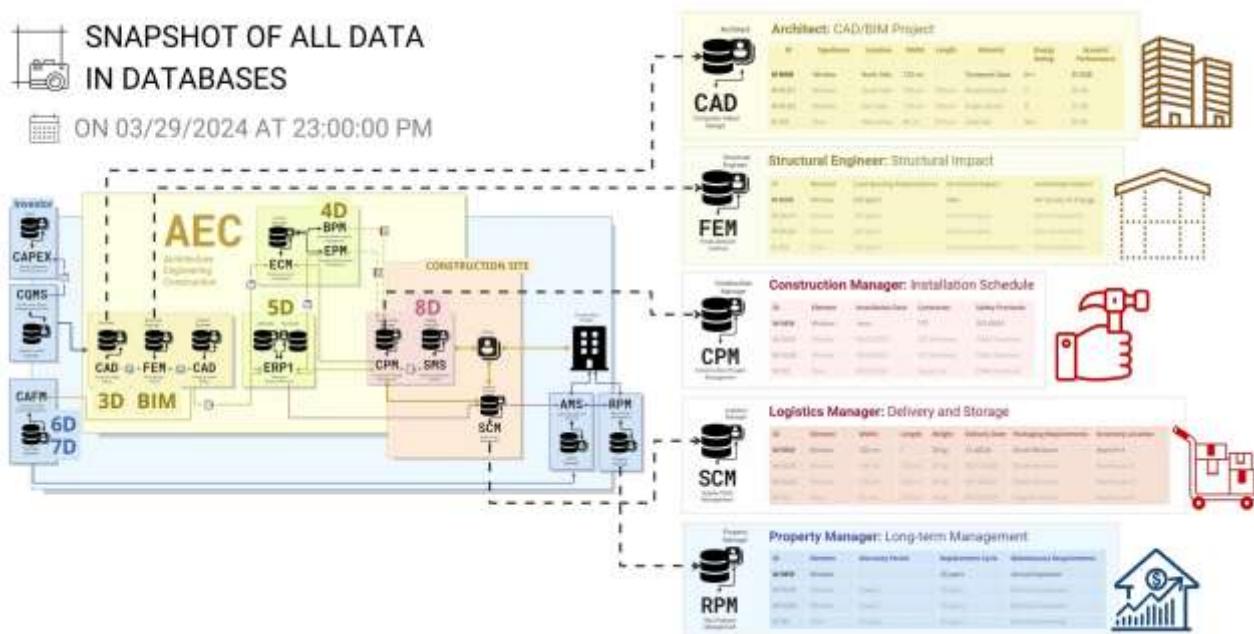
Obrázek 4.4-11 Strukturovaná data ze systémů CAD mohou mít podobu dvourozměrné tabulky se sloupci označujícími atributy prvků.

Skutečné projekty CAD (BIM) obsahují desítky nebo stovky tisíc prvků (obr. 9.1-10). Prvky v rámci formátů CAD jsou automaticky kategorizovány podle typu a kategorie - od oken a dveří až po desky, desky a stěny. Jedinečné identifikátory (např. nativní ID, které je automaticky nastaveno řešením CAD) nebo atributy typu (Název typu, Typ, Rodina) umožňují sledovat stejný objekt v různých systémech. Například nové okno na severní stěně budovy může být ve všech příslušných systémech organizace jednoznačně identifikováno jediným identifikátorem "W-NEW".

Zatímco názvy a identifikátory entit by měly být ve všech systémech konzistentní, soubor atributů a hodnot přiřazených těmto entitám se může v závislosti na kontextu použít výrazně lišit. Architekti, stavební inženýři, odborníci na výstavbu, logistiku a údržbu nemovitostí vnímají stejné prvky různým způsobem. Každý z nich se opírá o své vlastní klasifikátory, normy a cíle: některí se na okno dívají čistě z estetického hlediska a posuzují jeho tvar a proporce, zatímco jiní se na něj dívají z technického nebo provozního hlediska a analyzují tepelnou vodivost, způsob instalace, hmotnost nebo požadavky na údržbu. Při modelování dat a popisu prvků je proto důležité zohlednit všeobecnost jejich použití a zajistit konzistentnost dat při zohlednění oborových specifik.

Pro každou roli v podnikových procesech existují specializované databáze s vlastním uživatelským rozhraním - od návrhu a výpočtu až po logistiku, instalaci a provoz budovy (obr. 4.4-12). Každý takový systém je spravován profesionálním týmem specialistů prostřednictvím speciálního uživatelského rozhraní nebo prostřednictvím databázových dotazů, přičemž za souhrnem všech rozhodnutí učiněných na základě zadaných hodnot na konci řetězce stojí správce systému nebo vedoucí oddělení, který odpovídá za právní platnost a kvalitu zadaných údajů svým protistranám obsluhujícím ostatní

systémy.



Obrázek 4.4-12 Stejná entita má v různých systémech stejný identifikátor, ale různé atributy, které jsou důležité pouze v daném systému.

Jakmile jsme uspořádali sběr strukturovaných požadavků a dat na logické a fyzické úrovni, zbývá nám nastavit proces automatické kontroly dat z různých příchozích dokumentů a různých systémů s dříve shromážděnými požadavky.

Ověření údajů na a výsledky ověření

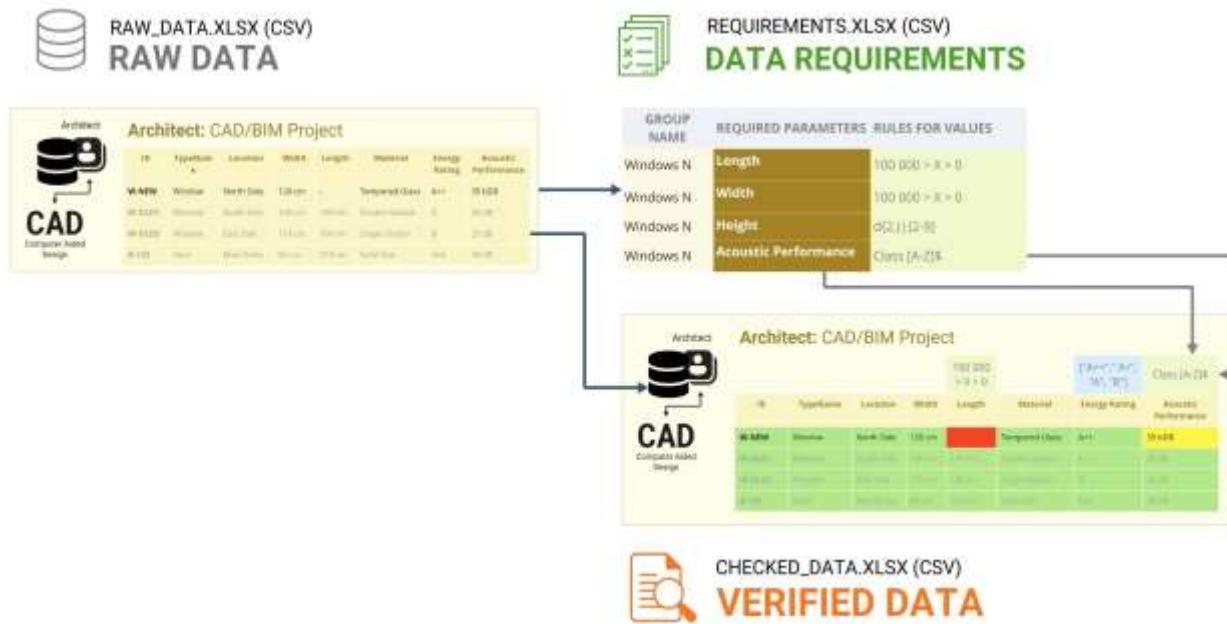
Všechna nová data vstupující do systému - ať už se jedná o dokumenty, tabulky nebo záznamy v databázi od klienta, architekta, inženýra, mistra, logistika nebo správce nemovitosti - musí být ověřena na základě dříve formulovaných požadavků (obr. 4.4-9). Proces validace je kritický: jakákoli chyba v datech může vést k nesprávným výpočtům, zpoždění harmonogramu, a dokonce k finančním ztrátám. Aby se tato rizika minimalizovala, měl by být zorganizován systematický a opakovatelný, iterativní postup validace dat.

Pro ověření nových dat vstupujících do systému - nestrukturovaných, textových nebo geometrických - je třeba je převést do volně strukturovaného nebo strukturovaného formátu. Proces validace pak musí data zkontolovat podle úplného seznamu požadovaných atributů a jejich povolených hodnot.

Převod různých typů dat: textu, obrázků, PDF -dokumentů a smíšených dat CAD (BIM) do strukturované podoby byl podrobně popsán v kapitole "Převod dat do strukturované podoby".

Příkladem je tabulka získaná z projektu CAD (BIM) (obr. 4.4-11). Obsahuje polostrukturovaná geometrická data a strukturované atributové informace o entitách projektu (obr. 3.1-14) - např. prvek z třídy "Windows".

Při validaci porovnáváme hodnoty atributů (obr. 4.4-11) s referenčními hraničními hodnotami, které byly definovány experty ve formě požadavku (obr. 4.4-9). Z výsledné srovnávací tabulky (obr. 4.4-13) bude zřejmé, které hodnoty jsou přijatelné a které je třeba před použitím dat mimo aplikace CAD (BIM) opravit.



Obrázek 4.4-13 V závěrečné validační tabulce jsou zvýrazněny ty hodnoty atributů nové entity třídy "Windows", kterým byste měli věnovat pozornost.

Při implementaci podobného řešení pomocí knihovny Pandas, kterou jsme popsali dříve v kapitole "Pandas: Nepochybatelný nástroj pro analýzu dat", budeme ověřovat data z tabulkového souboru získaného ze souboru CAD (BIM) (RVT, IFC, DWG, NWS, DGN) (obrázek 4.4-11) pomocí požadavků z jiného tabulkového souboru s požadavky (obrázek 4.4-9).

Abychom získali kód, musíme ve výzvě pro LLM popsat, že potřebujeme načíst data ze souboru **raw_data.xlsx** (kompletní soubor dat z databáze CAD (BIM)), **zkontrolovat** je a výsledek uložit do nového souboru **checked_data.xlsx** (obr. 4.4-13).

💡 Získejme kód pomocí LLM, aniž bychom se zmiňovali o knihovně Pandas:

Napište kód, který zkонтroluje tabulku ze souboru raw_data.xlsx a ověří je pomocí následujících validačních pravidel: hodnoty sloupců "Šířka" a "Délka" jsou větší než nula, "Energetické hodnocení" je zahrnuto v seznamu ["A++", "A+", "A", "B"] a "Akustický výkon" jako proměnná, kterou určíme později - s přidaným posledním validačním sloupcem, a uložte konečnou tabulku do nového souboru aplikace Excel checked_data.xlsx. ↵

- V odpovědi LLM bude popsán krátký příklad kódu v jazyce Python, který může být upřesněn a doplněn následnými výzvami :

The screenshot shows a user interface for generating code. At the top, there are four colored dots (red, yellow, green, blue) followed by the file name 'Validation.py'. To the right is a button labeled 'Create in LLM Chat'. Below this is a code editor window containing the following Python script:

```

1 import pandas as pd
2
3 # Load raw data from Excel file
4 df_data = pd.read_excel('raw_data.xlsx', sheet_name=0)
5
6 # Adding a 'Requirements Met' column that checks all conditions
7 df_data['Requirements Met'] = (
8     (df_data['Width'] > 0) &
9     (df_data['Length'] > 0) &
10    (df_data['Energy Rating'].isin(['A++', 'A+', 'A', 'B'])) &
11    (df_data['Acoustic Performance'] >= min_acoustic_performance)
12 )
13
14 # Save the results to a new Excel file
15 df_data.to_excel('checked_data.xlsx', index=False)

```

At the bottom of the code editor is a 'Run in IDE' button with icons for VS Code, PyCharm, and Jupyter Notebook.

Obrázek 4.4-14 Kód vygenerovaný modelem LLM- kontroluje převedený návrh CAD (BIM) podle požadavků na atributy ve formě hraničních hodnot.

Kód vygenerovaný jazykovým modelem LLM, lze použít v libovolném populárním IDE nebo online nástroji: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse s pluginem PyDev, Thonny, Wing IDE, IntelliJ IDEA s pluginem Python, JupyterLab nebo populární online nástroje Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Provedení kódu (obr. 4.4-14) ukáže, že "prvky entit" W-OLD1, W-OLD2, D-122 (a další prvky) z databáze CAD (BIM) splňují požadavky na atributy: šířka a délka jsou větší než nula a třída energetické účinnosti je jednou z hodnot seznamu "A++", "A", "B", "C" (obr. 4.4-15).

Potřebný a nedávno přidaný prvek W-NEW, který je zodpovědný za novou třídu prvků "Window" na

severní straně, není v souladu (atribut "Requirements Met"), protože jeho délka je nulová (hodnota "0.0" je podle našeho pravidla "Width">0 považována za nepřijatelnou) a neuvádí třídu energetické účinnosti.



CHECKED_DATA.XLSX (CSV)

VERIFIED DATA

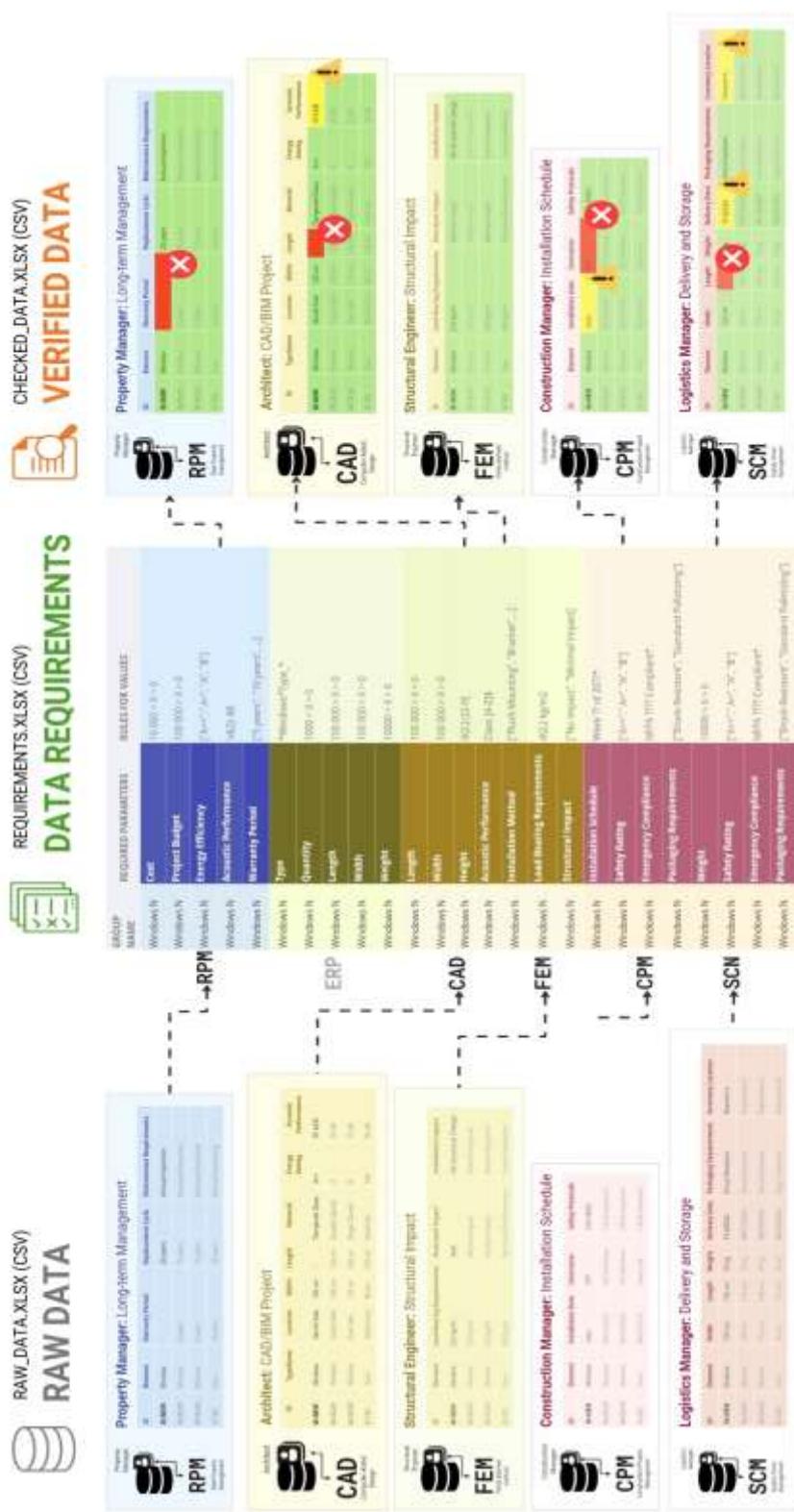
ID	TypeName	Location	Width	Length	Material	Energy Rating	Acoustic Performance	Requirements Met
0	W-NEW	Window	North Side	120	0.0	Tempered Glass		35
1	W-OLD1	Window	South Side	100	140.0	Double Glazed	A++	30
2	W-OLD2	Window	East Side	110	160.0	Single Glazed	B	25
3	D-122	Door	Main Entry	90	210.0	Solid Oak	B	30

Obrázek 4.4-15 Kontrola identifikuje entity, které neprošly procesem ověření, a přidá do výsledků nový atribut s hodnotami "False" nebo "True".

Podobně kontrolujeme konzistenci všech prvků projektu (entit) a požadovaných atributů pro každý ze systémů, tabulek nebo databází ve všech datech, která obdržíme od různých specialistů (obrázek 4.4-1) během procesu přidávání okna do projektu.

V závěrečné tabulce je vhodné pro vizualizaci barevně zvýraznit výsledky kontroly: zeleně jsou označeny atributy, které úspěšně prošly kontrolou, žlutě hodnoty s nekritickými odchylkami a červeně kritické neshody (obr. 4.4-16).

Výsledkem validace (obrázek 4.4-16) je seznam důvěryhodných a validovaných prvků s jejich identifikátory, u nichž bylo ověřeno, že splňují požadavky na atributy. Validované prvky poskytují jistotu, že tyto prvky splňují stanovené standardy a specifikace pro všechny systémy zapojené do procesu přidávání prvků třídy Okno nebo jakékoli jiné třídy (více o automatizaci validace dat a vytváření automatizovaného procesu ETL si povíme v kapitole "Automatizace ETL a validace dat").



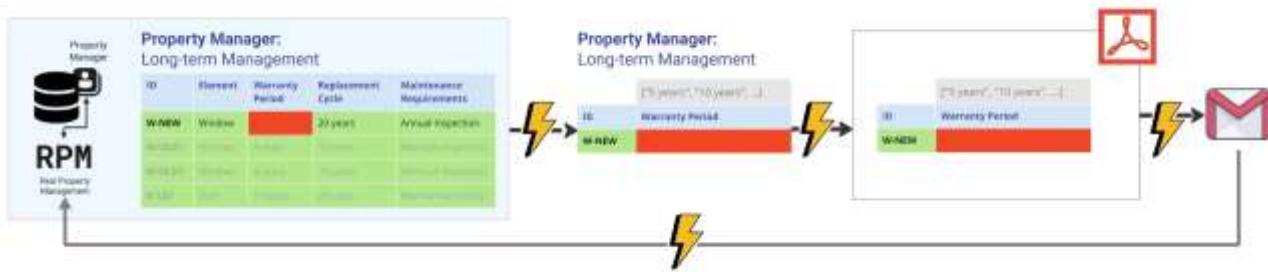
Obr. 4.4-16 Výsledek kontroly provedené pro všechny systémy umožňuje určit, které údaje nesplňují požadavky společnosti.

Subjektům, které byly úspěšně validovány, obvykle není třeba věnovat větší pozornost. Bez překážek postupují do dalších fází zpracování a integrace do jiných systémů. Na rozdíl od "kvalitních" položek jsou nejzajímavější ty, které validací neprošly. Informace o této odchylkách jsou klíčové: měly by být sdělovány nejen v tabulkových přehledech, ale také pomocí různých vizualizačních nástrojů. Grafické znázornění výsledků ověřování pomáhá rychle vyhodnotit celkový stav kvality dat, identifikovat problémové oblasti a neprodleně přijmout nápravná či korekční opatření.

Vizualizace výsledků ověřování

Vizualizace je základním nástrojem pro interpretaci výsledků kontroly. Kromě obvyklých souhrnných tabulek může obsahovat informační panely, diagramy a automaticky generované dokumenty ve formátu PDF, které seskupují prvky projektu podle jejich stavu kontroly. Barevné kódování zde může hrát podpůrnou roli: zelená barva může označovat prvky, které byly úspěšně ověřeny, žlutá barva může označovat prvky, které vyžadují další pozornost, a červená barva může označovat prvky, které mají kritické chyby nebo u nichž chybí klíčové údaje.

V našem příkladu (obr. 4.4-1) analyzujeme data z jednotlivých systémů krok za krokem: od CAD (BIM) a správy nemovitostí až po logistiku a harmonogram instalace (obr. 4.4-16). Po provedení auditu se pro každého specialistu automaticky vygenerují jednotlivá upozornění nebo dokumenty se zprávami, např. ve formátu PDF (obr. 4.4-17). Pokud jsou údaje správné, obdrží specialistka krátkou zprávu: "Děkujeme za spolupráci". Pokud jsou zjištěny nesrovnanosti, je zaslána podrobná zpráva ve znění: "Tento dokument obsahuje seznam prvků, jejich identifikátorů, atributů a hodnot, jejichž shoda nebyla zkontrolována".



Obrázek 4.4-17 Validace a automatické generování sestav urychluje proces hledání a pochopení nedostatků dat pro odborníka, který data vytváří.

Díky automatizovanému procesu validace - jakmile je zjištěna chyba nebo mezera v datech, je osobě odpovědné za vytvoření nebo zpracování příslušných entit a jejich atributů zasláno okamžité oznámení ve formě zprávy chatu, e-mailu nebo PDF -dokumentu (obr. 4.4-18) se seznamem prvků a popisů atributů, které nebyly validovány.



Obr. 4.4-18 Automatické kontrolní zprávy usnadňují pochopení chyb a urychlují práci při doplňování údajů o projektu.

Pokud například do systému správy majetku dorazí dokument (po strukturování) s nesprávně vyplněným atributem "Záruční doba", obdrží správce majetku upozornění se seznamem atributů, které je třeba zkontrolovat a opravit.

Stejně tak jakékoli nedostatky v plánu instalace nebo v logistických údajích povedou k automatickému vygenerování zprávy a například k zaslání oznámení v chatu nebo e-mailu s výsledky kontroly příslušnému specialistovi.

Kromě PDF -dokumentů a grafů s výsledky je možné vytvářet informační panely a interaktivní 3D - modely (obrázek 7.1-6, obrázek 7.2-12), které zvýrazňují prvky s chybějícími atributy, což uživatelům umožňuje vizuálně využívat 3D geometrie prvků k filtrování a hodnocení kvality a úplnosti těchto prvků v projektu.

Vizualizace výsledků inspekcí na v podobě automaticky generovaných dokumentů, grafů nebo řídicích panelů výrazně zjednoduší interpretaci dat a usnadňuje efektivní komunikaci mezi účastníky projektu.

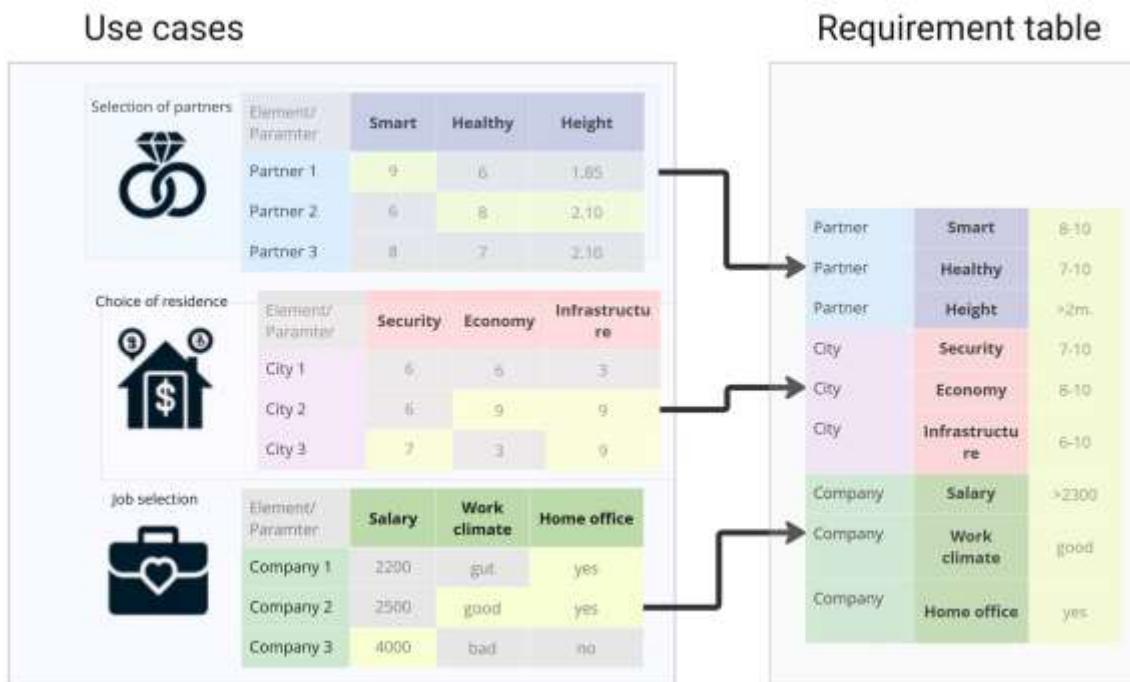
Proces automatické kontroly dat z různých systémů a informačních zdrojů lze přirovnat k informovanému rozhodování v každodenním životě. Stejně jako firmy ve stavebnictví zvažují mnoho proměnných - od spolehlivosti vstupních dat až po jejich vliv na čas, náklady a kvalitu realizace projektu -, tak i člověk při důležitých rozhodnutích, například při výběru místa k bydlení, zvažuje celou řadu faktorů: dopravní dostupnost, infrastrukturu, náklady, bezpečnost, kvalitu života. Všechna tato hlediska vytvářejí systém kritérií, na jejichž základě jsou činěna konečná rozhodnutí, z nichž se skládá náš život.

Srovnání kontrol kvality dat s potřebami lidského života

Navzdory neustálému vývoji metod a nástrojů pro kontrolu kvality dat zůstává základní princip souladu informací nezměněn. Tento princip je zabudován do základů vyspělého systému řízení, ať už v podnikání, nebo v každodenním životě.

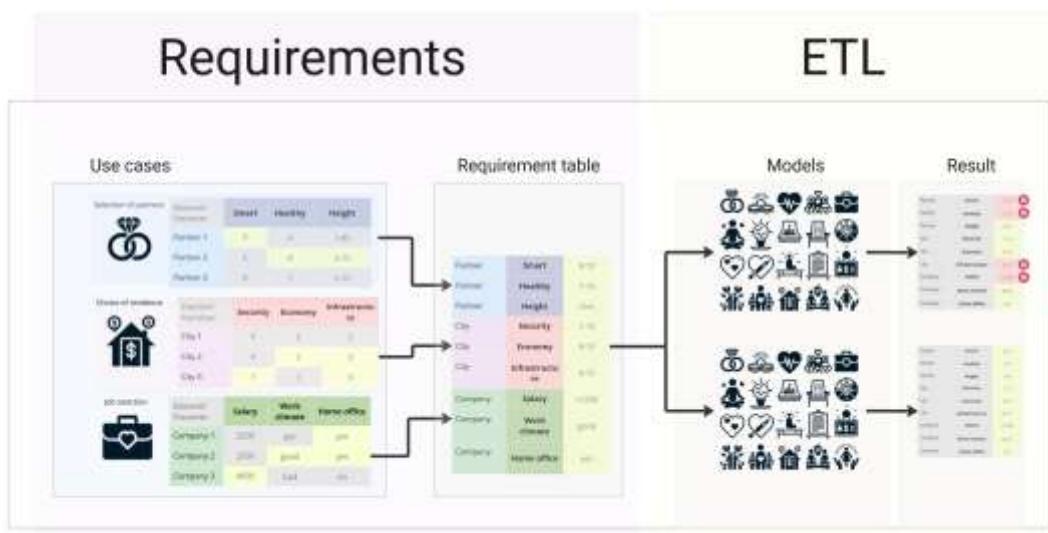
Proces opakovaného ověřování dat se podobá rozhodovacímu procesu, kterému každý člověk denně čelí. V obou případech se spolehlíme na zkušenosti, data a nové informace, které jsou k dispozici. A stále více životních a profesních rozhodnutí - od strategických až po každodenní rozhodování - se činí na základě dat.

Například při výběru bydlení nebo životního partnera si v mysli intuitivně vytváříme tabulku kritérií a vlastností, podle kterých porovnáváme alternativy (obr. 4.4-19). Tyto charakteristiky - ať už jde o osobní vlastnosti člověka nebo parametry nemovitosti - představují atributy, které ovlivňují konečné rozhodnutí.



Obr. 4.4-19 Volba bydliště, zaměstnání nebo partnerství vychází z individuálních požadavků na atributy.

Používání strukturovaných údajů a formalizovaný přístup k popisu požadavků (obrázek 4.4-20) přispívá k informovanějším a informovaným rozhodnutím v profesním i osobním životě.

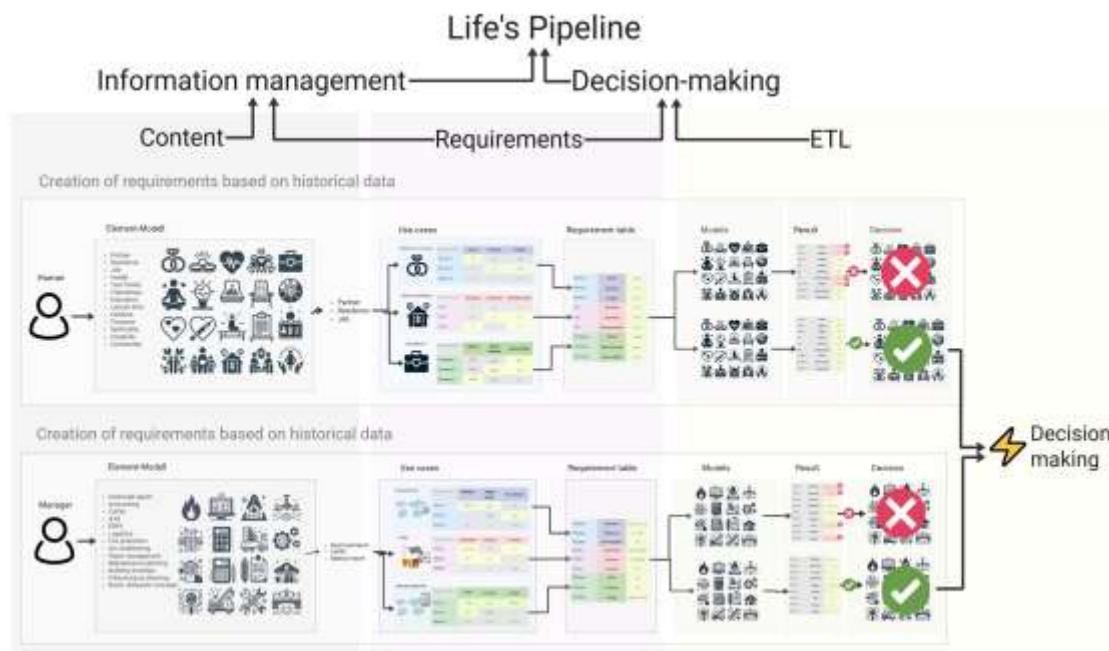


Obr. 4.4-20 Formalizace požadavků umožňuje systematizovat vnímání životních a obchodních rozhodnutí.

Rozhodování založené na datech není výhradně obchodním nástrojem. Bez problémů se začleňuje i do běžného života, a to podle běžných kroků zpracování dat (obr. 4.4-21) podobně jako proces ETL (Extract, Transform, Load), o kterém jsme již hovořili na začátku této části při strukturování dat a který podrobně prozkoumáme v souvislosti s automatizací úloh v sedmé části knihy:

- **Data jako základ (výpis):** V každé oblasti - ať už jde o práci nebo osobní život - shromažďujeme informace. V podnikání to mohou být zprávy, čísla, údaje o trhu; v osobním životě to mohou být osobní zkušenosti, rady blízkých, zpětná vazba, postřehy.
- **Kritéria hodnocení (transformace):** shromázděné informace se interpretují na základě předem stanovených kritérií. V práci jsou to ukazatele výkonnosti (KPI), rozpočtová omezení a normy; v osobním životě parametry jako cena, pohodlí, spolehlivost, charisma atd.
- **Předpověď a analýza rizik (Load):** poslední fáze zahrnuje rozhodování na základě analýzy transformovaných dat a porovnání možných důsledků. Je to podobné jako u obchodních procesů, kde data procházejí obchodní logikou a filtrem rizik.

Rozhodnutí, která děláme - od triviálních preferencí, jako je výběr jídla k snídani, až po zásadní životní události, jako je volba povolání nebo životního partnera - jsou ze své podstaty výsledkem zpracování a vyhodnocování dat.



Obr. 4.4-21 Podnikání a život obecně je řadou rozhodnutí založených na datech, přičemž kvalita dat použitých k rozhodování je klíčovým faktorem.

Vše v našem životě je vzájemně propojeno a stejně jako se živé organismy, včetně člověka, řídí přírodními zákony, vyvíjejí se a přizpůsobují měnícím se podmínkám, tak i lidské procesy, včetně toho, jak shromažďujeme a analyzujeme data, odrázejí tyto přírodní principy. Úzký vztah mezi přírodou a lidskou činností potvrzuje nejen naši závislost na přírodě, ale také naši touhu uplatňovat zákony zdokonalené miliony let evoluce při vytváření datových architektur, procesů a systémů pro rozhodování.

Nové technologie, zejména ve stavebnictví, jsou ukázkovým příkladem toho, jak se lidstvo stále inspiruje přírodou, aby vytvářelo lepší, udržitelnější a efektivnější řešení.

Další kroky: proměna dat v přesné výpočty a plány

V této části jsme se zabývali tím, jak převést nestrukturovaná data do strukturovaného formátu, vytvořit datové modely a organizovat procesy pro kontrolu kvality informací ve stavebních projektech. Správa, standardizace a klasifikace dat je zásadní proces, který vyžaduje systematický přístup a jasné pochopení obchodních požadavků. Techniky a nástroje probírané v této části umožňují spolehlivou integraci mezi různými systémy v průběhu celého životního cyklu objektu.

Shrňme tuto část a zdůrazněme hlavní praktické kroky, které vám pomohou aplikovat diskutované přístupy při plnění každodenních úkolů:

- Začněte systematizací požadavků
 - Vytvoření registru atributů a parametrů pro klíčové prvky vašich projektů a procesů.
 - Zdokumentujte hraniční hodnoty pro každý atribut

- Vizualizovat procesy a vztahy mezi třídami, systémy a atributy pomocí vývojových diagramů (např. v programech Miro, Canva, Visio).

■ Automatizace převodu dat

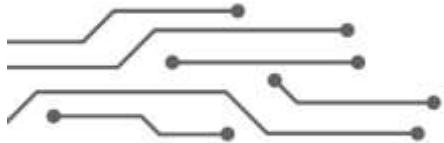
- Zjistěte, které z vašich dokumentů, které se často používají v procesech, lze digitalizovat pomocí knihoven OCR a převést do tabulkové podoby.
- Podívejte se na nástroje pro reverzní inženýrství pro získávání dat z CAD (BIM).
- Zkuste nastavit automatické načítání dat z dokumentů nebo formátů, které často používáte při své práci, do tabulkového formuláře.
- Nastavení automatických konverzí mezi různými datovými formáty

■ Vytvoření znalostní báze pro kategorizaci

- Vyvinout interní nebo použít existující klasifikátor prvků v souladu s průmyslovými standardy.
- dokumentovat vzájemné vztahy mezi různými klasifikačními systémy.
- Diskutujte se svým týmem na téma používání jednotného systému identifikace a jednoznačné klasifikace prvků.
- Začněte budovat proces automatické validace dat - jak těch, se kterými pracujete v rámci týmu, tak těch, která jsou předávána do externích systémů.

Pomocí těchto přístupů můžete výrazně zlepšit kvalitu svých dat a zjednodušit jejich následné zpracování a transformaci. V následujících částech této knihy se budeme zabývat tím, jak již strukturovaná a připravená data použít pro automatizované výpočty, odhady nákladů, plánování a řízení stavebních projektů.





V ČÁST

KALKULACE NÁKLADŮ A ČASU: ZAČLENĚNÍ DAT DO STAVEBNÍCH PROCESŮ.

Pátá část je věnována praktickým aspektům využití dat pro optimalizaci kalkulace nákladů a plánování stavebních projektů. Podrobně je rozebrána metoda odhadu nákladů založená na zdrojích a automatizace procesů odhadu. Zvažují se metody automatizovaného získávání Quantity Take-Off) z CAD (BIM)-modelů a jejich integrace s kalkulačními systémy. Jsou zkoumány techniky 4D a 5D modelování pro časové plánování a řízení stavebních nákladů s konkrétními příklady jejich použití. Představena je analýza rozšířených informačních vrstev 6D -8D, které poskytují integrovaný přístup k posuzování udržitelnosti, provozu a bezpečnosti nemovitých objektů. Podrobně jsou rozebrány metody výpočtu uhlíkové stopy a ESG -indikátorů stavebních projektů v kontextu moderních environmentálních požadavků a norem. Kriticky jsou zhodnoceny možnosti a omezení tradičních systémů ERP a PMIS v řízení stavebních procesů s analýzou jejich dopadu na transparentnost cen. Jsou předpovězeny vyhlídky na přechod od uzavřených řešení k otevřeným standardům a flexibilním nástrojům pro analýzu dat, které jsou schopny zajistit vyšší efektivitu stavebních procesů.

KAPITOLA 5.1.

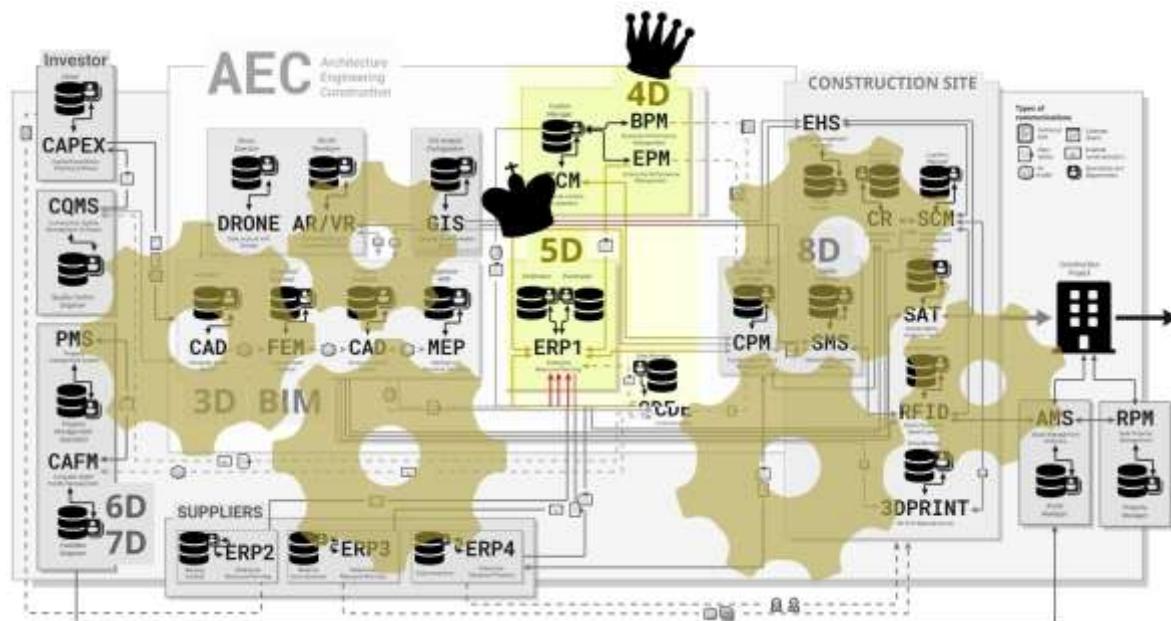
VÝPOČTY A ODHADY NÁKLADŮ NA STAVEBNÍ PROJEKTY

Základy stavebnictví: odhad množství, nákladů a času

Mezi mnoha podnikovými procesy, které rozhodují o udržitelnosti podniku ve stavebnictví, mají zvláštní význam - stejně jako před tisíci lety - procesy přesného odhadu počtu prvků, nákladů na projekt a dodací lhůty (obr. 5.1-1).

Rozvoj písma byl výsledkem řady faktorů, včetně potřeby zaznamenávat hospodářské transakce, obchod a hospodaření se zdroji v raných společnostech. První právně významné dokumenty, hliněné tabulky s výpočty nákladů na materiál a mzdy, se používaly v souvislosti s obchodem a stavebnictvím. Tyto tabulky zaznamenávaly závazky stran při výstavbě staveb a byly uchovávány jako důkaz dohod a peněžních a obchodních vztahů.

Po tisíce let zůstával přístup k odhadům v podstatě nezměněn: výpočty se prováděly ručně a spoléhaly se na zkušenosť a intuici odhadujícího inženýra. S nástupem modulárních systémů ERP a nástrojů CAD se však tradiční přístup k odhadování množství, nákladů a času začal rychle proměňovat. Dnešní digitální technologie umožňují plně automatizovat klíčové výpočty času a nákladů, což umožnuje větší přesnost, rychlosť a transparentnost při plánování zdrojů stavebních projektů.



Obr. 5.1-1 Z mnoha různých systémů jsou v podnikání nejdůležitější nástroje zodpovědné za měření objemu, nákladů a času.

Stavební firmy se zaměřují na přesné údaje o čase a nákladech. Ty zase závisí na množství použitého materiálu a práce a jejich transparentnost ovlivňuje ziskovost. Složitost kalkulačních procesů a jejich netransparentnost však často vedou k vyšším nákladům na projekt, nedodržení termínů a dokonce k úpadku.

Podle zprávy KPMG "Známé problémy - nové přístupy" (2023) je pouze 50% stavebních projektů dokončeno včas a 87% společností uvádí zvýšenou kontrolu ekonomiky investičních projektů. Hlavní problémy souvisejí s nedostatkem kvalifikovaného personálu a obtížným předvídáním rizik [2].

Historické údaje o kalkulaci nákladů a procesních časech se shromažďují během výstavby minulých projektů po celou dobu existence stavební společnosti a vkládají se do databází různých systémů (ERP, PMIS BPM, EPM atd.).

Kvalitní historické údaje o nákladech jsou pro stavební organizaci velkou konkurenční výhodou, která přímo ovlivňuje její schopnost přežít.

Oddělení odhadu a kalkulace nákladů ve stavebních a inženýrských společnostech jsou vytvořena za účelem shromažďování, ukládání a aktualizace historických dat o kalkulačních projektů. Jejich hlavním úkolem je shromažďovat a systematizovat zkušenosti společnosti, což umožnuje v průběhu času zlepšovat přesnost odhadu rozsahu, harmonogramu a nákladů nových projektů. Tento přístup pomáhá minimalizovat chyby v budoucích výpočtech na základě praxe a výsledků již realizovaných projektů.

Metody výpočtu odhadovaných nákladů projektů

Specialisté na kalkulaci nákladů používají různé metody odhadu, přičemž každá z nich se zaměřuje na určitý typ údajů, dostupnost informací a úroveň podrobnosti projektu. Mezi nejběžnější patří:

- **Metoda založená na zdrojích:** odhad předpokládaných nákladů projektu na základě podrobné analýzy všech potřebných zdrojů, jako jsou materiály, vybavení a pracovní síla. Tato metoda vyžaduje podrobný seznam všech úkolů a zdrojů potřebných k provedení každého úkolu a následný výpočet jejich nákladů. Tato metoda je velmi přesná a při odhadu nákladů se hojně využívá.
- **Parametrická metoda:** používá statistické modely k odhadu nákladů na základě parametrů projektu. Může zahrnovat analýzu nákladů na měrnou jednotku, jako je plocha budovy nebo rozsah prací, a přizpůsobení těchto hodnot konkrétním podmínkám projektu. Metoda je účinná zejména v počátečních fázích, kdy ještě nejsou k dispozici podrobné informace.
- **Jednotková metoda (metoda jednotkových nákladů):** vypočítá odhadované náklady projektu na základě nákladů na měrnou jednotku (např. na metr čtvereční nebo metr krychlový). To umožnuje rychlý a snadný způsob porovnání a analýzy nákladů různých projektů nebo jejich částí.
- **Expertní posudek (metoda Delphi):** vychází z názorů expertů, kteří na základě svých zkušeností a znalostí odhadují hodnotu projektu. Tento přístup je užitečný, pokud nejsou k dispozici přesné výchozí údaje nebo pokud je projekt jedinečný.

Je třeba poznamenat, že parametrickou metodu a expertní posudky lze přizpůsobit modelům strojového učení. To umožňuje automatické generování prognóz nákladů a harmonogramu projektu na základě trénovacích vzorků. Příklady použití takových modelů jsou podrobněji rozebrány v kapitole "Příklad použití strojového učení pro zjištění nákladů a harmonogramů projektu" (obr. 9.3-5).

Přesto zůstává metoda založená na zdrojích nejoblíbenější a nejrozšířenější ve světové praxi. Poskytuje nejen přesné posouzení odhadovaných nákladů, ale umožňuje také vypočítat dobu trvání jednotlivých procesů na staveništi i celého projektu jako celku (podrobněji v kapitole "Harmonogramy výstavby a 4D -projektová data").

Metoda založená na zdrojích kalkulace nákladů a odhadu ve stavebnictví.

Kalkulace nákladů založená na zdrojích je metoda manažerského účetnictví, při níž jsou náklady na projekt založeny na přímém vyúčtování všech zapojených zdrojů. Ve stavebnictví tento přístup zahrnuje podrobnou analýzu a vyhodnocení všech materiálových, pracovních a technických zdrojů potřebných k provedení díla.

Metoda založená na zdrojích, poskytuje vysokou míru transparentnosti a přesnosti při plánování rozpočtu, protože se zaměřuje na skutečné ceny zdrojů v době odhadu. To je důležité zejména v nestabilním ekonomickém prostředí, kde mohou výkyvy cen výrazně ovlivnit celkové náklady projektu.

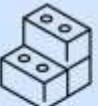
V následujících kapitolách se budeme podrobně zabývat procesem kalkulace nákladů na základě zdrojů. Abychom lépe pochopili jeho principy ve stavebnictví, použijeme analogii s kalkulací nákladů na večeři v restauraci. Vedoucí restaurace při plánování večera sestaví seznam potřebných produktů, zohlední dobu přípravy jednotlivých pokrmů a poté náklady vynásobí počtem hostů. Ve stavebnictví je postup podobný: pro každou kategorii prvků projektu (objektů) se vytvoří položkové odhadu Recepty, a celkové náklady projektu se určí součtem všech nákladů v celkovém účtu - konečném odhadu podle kategorií.

Klíčovou a počáteční fází přístupu založeného na zdrojích je vytvoření počáteční databáze společnosti. V první fázi kalkulace nákladů je sestaven strukturovaný seznam všech položek, materiálů, druhů prací a zdrojů, které má společnost v rámci svých stavebních projektů k dispozici - od hřebíku ve skladu až po popis lidí přes jejich kvalifikaci a hodinovou sazbu. Tyto informace jsou systematizovány do jednotné "Databáze stavebních zdrojů a materiálů" - tabulkového registru obsahujícího údaje o názvech, charakteristikách, měrných jednotkách a aktuálních cenách. Právě tato databáze se stává hlavním a primárním zdrojem informací pro všechny následné výpočty zdrojů - nákladů i časového rozvržení prací.

Databáze stavebních zdrojů: katalog stavebních materiálů a prací

Databáze nebo tabulka stavebních zdrojů a materiálů - obsahuje podrobné informace o každém prvku, který lze použít ve stavebním projektu - výrobku, položce, materiálu nebo službě, včetně jeho názvu,

popisu, měrné jednotky a jednotkové ceny, zaznamenané ve strukturované podobě. V této tabulce najdete vše od různých druhů paliv a materiálů používaných v projektech až po podrobné seznamy specialistů v podobě různých kategorií s popisem hodinových sazeb (obr. 5.1-2).

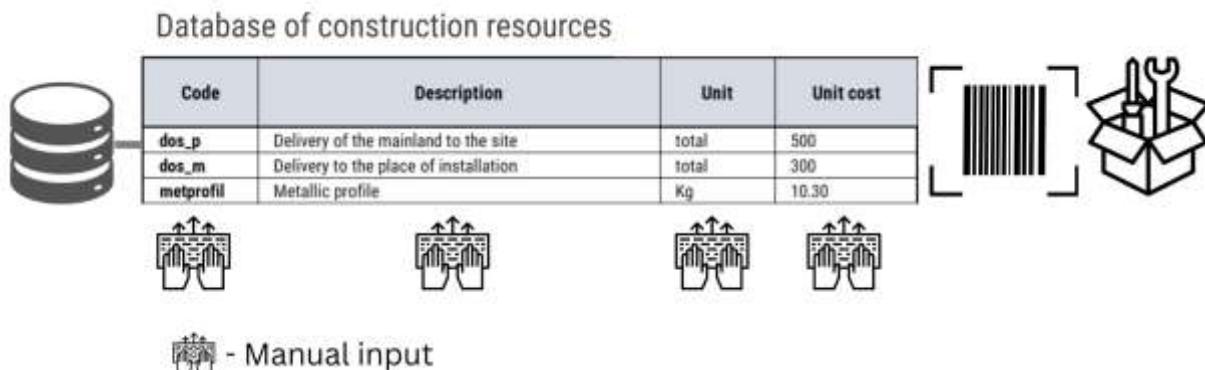
Database of resources			
 1st grade potatoes 1 kg \$2,99	Sand lime bricks	 1 pcs \$1	JCB 3CX backhoe loader
	Black Angus marble beef		 1 h \$150
	Broccoli		Laborer of the 1st category  1 h \$30

Obrázek 5.1-2 Tabulka zdrojů je seznam složek popisující materiál a službu s jednotkovými náklady.

"Databáze zdrojů" se podobá katalogu produktů v internetovém obchodě, kde je u každé položky uveden podrobný popis jejích vlastností. To usnadňuje kvantitativním odhadcům výběr správných zdrojů (podobně jako výběr výrobků při přidávání do nákupního košíku) potřebných pro výpočet konkrétních stavebních procesů formou kalkulací (konečná objednávka v internetovém obchodě).

Databázi zdrojů si lze představit také jako seznam všech ingrediencí v kuchařské knize restaurace. Každý stavební materiál, vybavení a služba jsou podobné ingrediencím používaným v receptech. "Databáze zdrojů" je podrobný seznam všech ingrediencí - stavebních materiálů a služeb, včetně jejich ceny za jednotku: kus, metr, hodinu, litr atd.

Nové prvky entit lze do tabulky "Databáze stavebních zdrojů" přidávat dvěma způsoby - ručně (obr. 5.1-3) nebo automaticky integrací se systémy řízení zásob nebo databázemi dodavatelů společnosti.



Obr. 5.1-3 Databáze zdrojů se vyplňuje ručně nebo automaticky přebírá údaje z jiných databází.

Typická středně velká stavební firma využívá databázi obsahující tisíce a někdy i desetitisíce položek s podrobným popisem, které lze využít ve stavebních projektech. Tato data jsou pak automaticky využívána ve smlouvách a projektové dokumentaci k přesnému popisu skladby prací a procesů

Aby bylo možné držet krok s měnícími se podmínkami na trhu, jako je inflace, je atribut "jednotkové náklady" pro každý produkt (zboží nebo službu) v databázi zdrojů (obrázek 5.1-3) pravidelně aktualizován ručně nebo automatickým stahováním aktuálních cen z jiných systémů nebo online platforem.

Aktualizace jednotkových nákladů na zdroj může být prováděna měsíčně, čtvrtletně nebo ročně - v závislosti na povaze zdroje, inflaci a vnějším ekonomickém klimatu. Tyto aktualizace jsou nezbytné pro zachování přesnosti výpočtů a odhadů, neboť tyto základní prvky jsou výchozím bodem pro práci odhadců nákladů. Aktuální údaje slouží k tvorbě odhadů, rozpočtů a harmonogramů, které odrážejí skutečné podmínky na trhu a snižují riziko chyb v následních výpočtech projektu.

Sestavení kalkulací a vyčíslení nákladů na práce na základě zdrojové základny.

Po naplnění "Databáze stavebních zdrojů" (obr. 5.1-3) minimálními jednotkovými entitami můžete začít vytvářet kalkulace, které se počítají pro každý proces nebo práci na staveništi pro určité měrné jednotky: například pro jeden metr krychlový betonu, jeden metr čtvereční sádrokartonové stěny, pro metr obrubníku nebo pro montáž oken.

Například pro stavbu cihlové zdi o rozloze 1 m² (obrázek 5.1-4) je na základě zkušeností z předchozích projektů zapotřebí přibližně 65 cihel (entita "Silikátová cihla") v ceně 1 USD za kus (atribut "Náklady na kus"), celkem 65 USD. Podle mých zkušeností je také nutné použít stavební zařízení (entita "JCB 3CX Loader") na 10 minut, které umístí cihly v blízkosti pracovního prostoru. Vzhledem k tomu, že pronájem zařízení stojí 150 USD za hodinu, 6 minut použití by stálo přibližně 15 USD. Kromě toho bude zapotřebí na 2 hodiny zajistit dodavatele pro pokládku cihel, jehož hodinová sazba bude 30 USD a celková částka 60 USD.

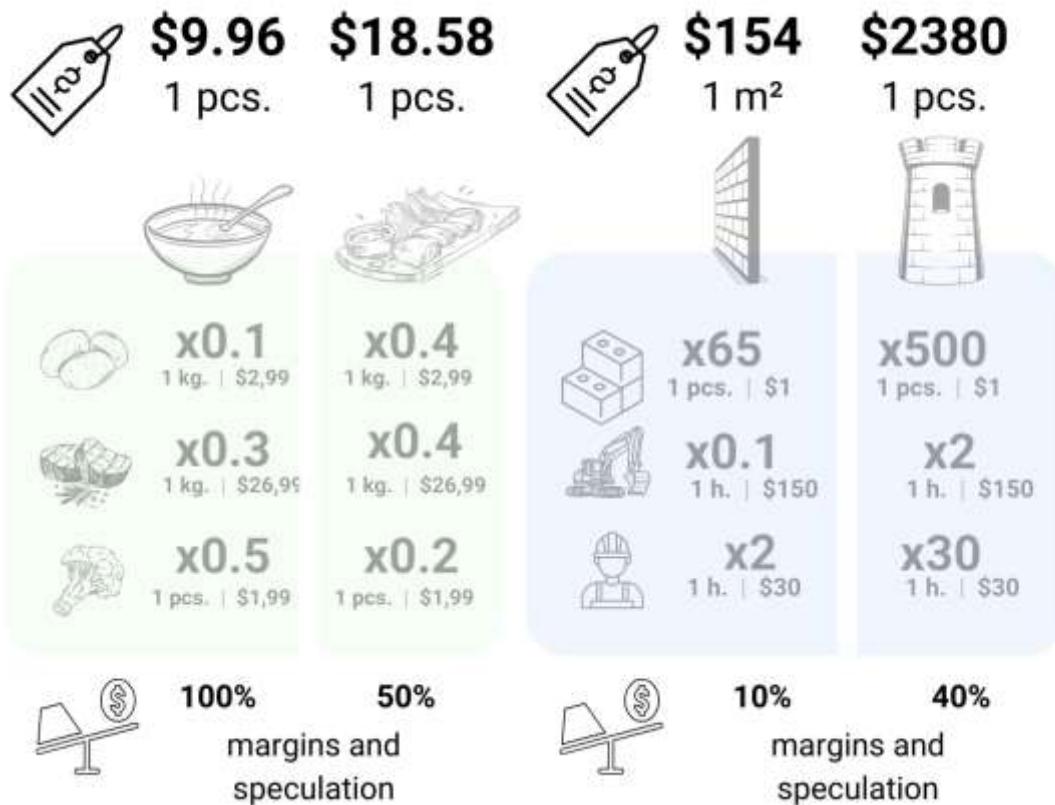
x0.1 1 kg \$2,99	x0.4 1 kg. \$2,99	x65 1 pcs \$1	x500 1 pcs \$1
x0.3 1 kg \$26,99	x0.4 1 kg. \$26,99	x0.1 1 h \$150	x2 1 h \$150
x0.5 1 pcs \$1,99	x0.2 1 pcs. \$1,99	x2 1 h \$30	x30 1 h \$30
+ margins and speculation	+ margins and speculation	+ margins and speculation	+ margins and speculation

Obr. 5.1-4 Kalkulace nákladů obsahuje podrobný výčet stavebních materiálů a služeb potřebných k provedení prací a procesů.

Složení kalkulací (tzv. "receptur") se vytváří na základě historických zkušeností, které společnost nashromázdila při provádění velkého objemu podobných prací. Tyto praktické zkušenosti jsou obvykle nashromázděny na základě zpětné vazby ze stavby. Zejména mistr sbírá informace přímo na stavbě, zaznamenává skutečné mzdrové náklady, spotřebu materiálu a nuance technologických operací. Ve spolupráci s oddělením odhadu se pak tyto informace iterativně zpřesňují: zpřesňují se popisy postupů, upravuje se skladba zdrojů a aktualizují se kalkulace nákladů podle skutečných údajů z posledních projektů.

Stejně jako recept popisuje ingredience a množství potřebné k přípravě pokrmu, odhad nákladů poskytuje podrobný seznam všech stavebních materiálů, zdrojů a služeb potřebných k dokončení určité práce nebo procesu.

Pravidelně prováděné práce umožňují pracovníkům, mistrům a odhadcům orientovat se v potřebném množství zdrojů: materiálu, paliva, pracovního času a dalších parametrů potřebných k provedení jednotky práce (obr. 5.1-5). Tyto údaje se do odhadních systémů zadávají ve formě tabulek, kde je každý úkol a operace popsána prostřednictvím minimálních prvků zdrojové základny (s neustále aktualizovanými cenami), což zajišťuje přesnost výpočtů.



Obr. 5.1-5 Pro každou zakázku se zjišťují jednotkové sazby, kdy se atribut objemu subjektu vynásobí jeho množstvím a přičte se procento zisku.

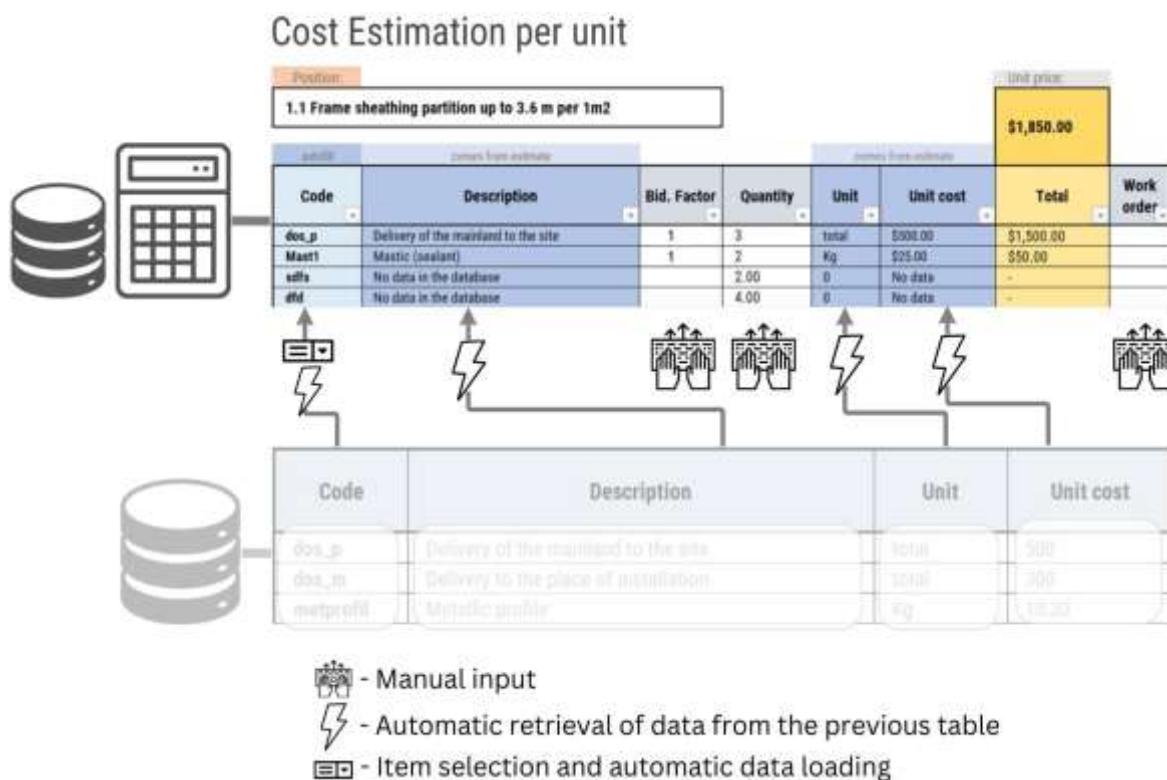
Pro získání celkových nákladů každého procesu nebo činnosti (objektu kalkulace nákladů) se atribut nákladů vynásobí jeho počtem a faktory. Koeficienty mohou zohledňovat různé faktory, například složitost práce, regionální charakteristiky, míru inflace, potenciální rizika (očekávané procento režie) nebo spekulaci (faktor dodatečného zisku).

Odhadce jako analytik převádí zkušenosti a doporučení mistra do standardizovaných odhadů, které popisují stavební procesy prostřednictvím zdrojových entit v tabulkové formě. Úkolem odhadce je v podstatě shromažďovat a strukturovat prostřednictvím parametrů a koeficientů informace přicházející ze stavby.

Konečné náklady na jednotku práce (např. metr čtvereční nebo krychlový nebo jedna instalace jedné jednotky) tak zahrnují nejen přímé náklady na materiál a práci, ale také firemní přírážky, režijní náklady, pojištění a další faktory (obrázek 5.1-6)

Zároveň se již nemusíme starat o skutečné ceny při výpočtech (receptů), protože skutečné ceny jsou vždy zohledněny v "základně zdrojů" (tabulce ingrediencí). Na úrovni výpočtů se do tabulky automaticky načítají údaje ze surovinové základny (např. podle kódu položky nebo jejího jedinečného identifikátoru), které načítají popis a skutečnou cenu za jednotku, kterou lze zase automaticky načíst

z online platformem nebo online obchodu se stavebními materiály. Odhadce na úrovni kalkulací prací musí pouze popsat práci nebo proces prostřednictvím atributu "množství zdrojů" a dalších faktorů.



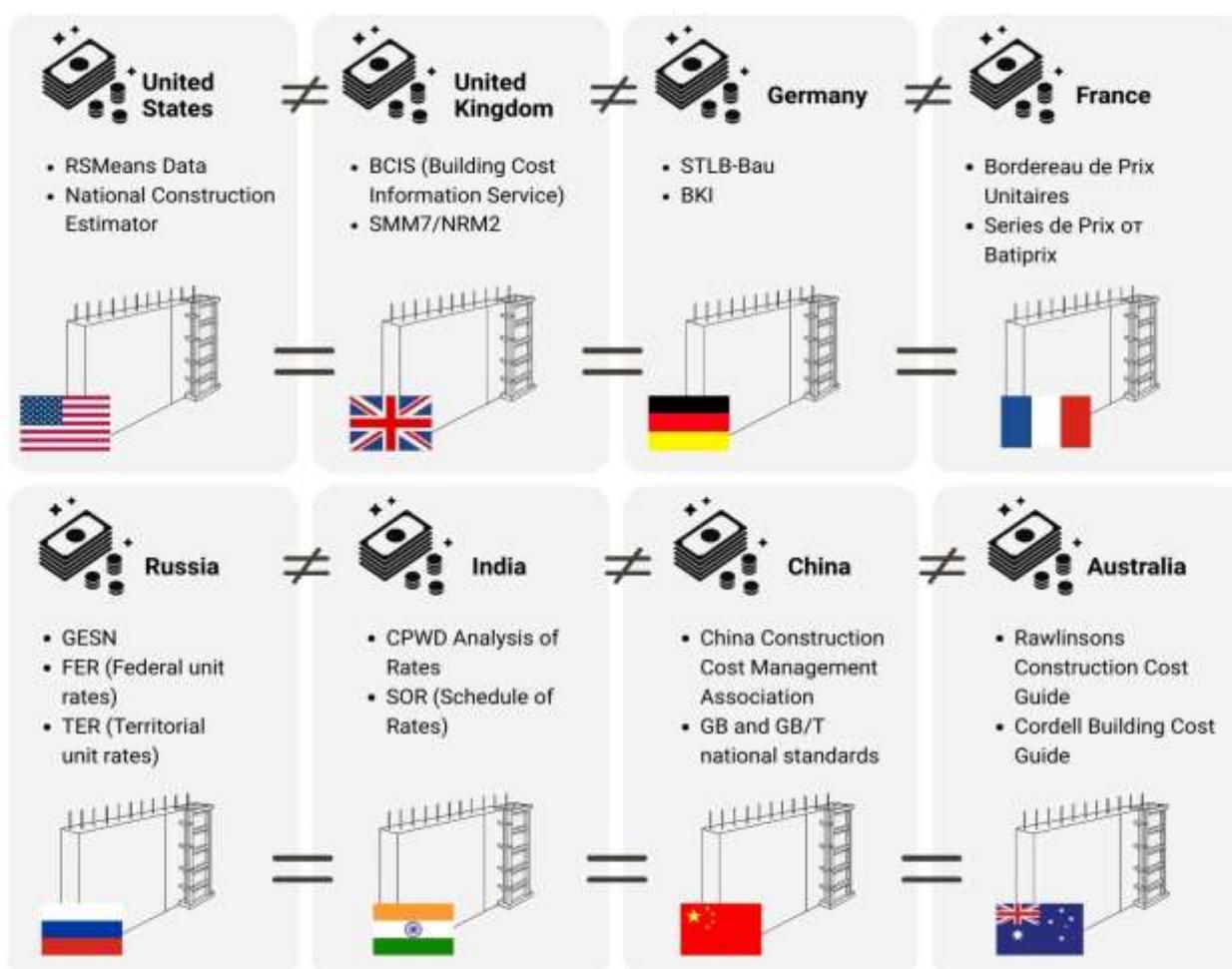
Obr. 5.1-6 Ve fázi výpočtu jednotkových nákladů práce se vyplňují pouze atributy množství potřebných zdrojů, vše ostatní se automaticky načte z databáze zdrojů.

Vytvořené kalkulace zakázek jsou uloženy ve formě vzorových tabulek typových projektů, které jsou přímo propojeny s databází stavebních zdrojů a materiálů. Tyto šablony představují standardizované předpisy pro opakující se typy prací pro budoucí projekty, čímž je zajištěna jednotnost kalkulací v celé společnosti.

Pokud se v databázi změní náklady na jakýkoli zdroj (obr. 5.1-3) - ať už ručně nebo automaticky prostřednictvím stahování aktuálních tržních cen (např. v podmírkách inflace) - aktualizace se okamžitě promítne do všech propojených kalkulací (obr. 5.1-6). To znamená, že je třeba měnit pouze zdrojovou základnu, zatímco šablony kalkulací a odhady nákladů zůstávají v čase nezměněny. Tento přístup zajišťuje stabilitu a reprodukovatelnost výpočtů pro případné cenové výkyvy, které jsou zohledněny pouze v relativně jednoduché tabulce zdrojů (obr. 5.1-3).

Pro každý nový projekt se vytvoří kopie standardní šablony kalkulace nákladů, což umožnuje provádět změny a upravovat činnosti podle konkrétních požadavků, aniž by se měnila původní šablona přijatá společností. Tento přístup poskytuje flexibilitu při úpravě kalkulací: můžete zohlednit specifika staveniště, přání zákazníka, zavést koeficienty rizika nebo ziskovosti (spekulace) - to vše bez porušení firemních standardů. To firmě pomáhá najít rovnováhu mezi maximalizací zisku, spokojeností zákazníka a zachováním konkurenčeschopnosti.

V některých zemích jsou takovéto šablony kalkulace nákladů, které se hromadí po desetiletí, standardizovány na národní úrovni a stávají se součástí standardů národního systému kalkulace nákladů ve stavebnictví (obrázek 5.1-7).



Obr. 5.1-7 Různé země světa mají pro kalkulaci nákladů na stejný prvek vlastní pravidla s vlastními (normativními) kompendii a normami pro stavební práce.

Takové standardizované základny odhadů zdrojů (obrázek 5.1-7) jsou povinné pro použití všemi účastníky trhuzejměna pro projekty financované z veřejných zdrojů. Taková standardizace zajišťuje

transparentnost, srovnatelnost a spravedlnost při tvorbě cen a smluvních závazků pro klienta

Konečné vyčíslení nákladů projektu: od odhadů k rozpočtům

Státní a oborové normy pro odhady hrají ve stavební praxi v různých zemích různou roli. Zatímco některé země vyžadují striktní dodržování jediné normy, většina vyspělých ekonomik uplatňuje flexibilnejší přístup. V tržních ekonomikách slouží státní stavební normy obvykle pouze jako výchozí. Stavební společnosti tyto normy přizpůsobují svým provozním modelům nebo je zcela přepracovávají a doplňují o vlastní přizpůsobené faktory. Tyto úpravy odrážejí firemní zkušenosti, efektivitu hospodaření se zdroji a často i faktory, v nichž mohou být zohledněny například spekulativní zisky společnosti.

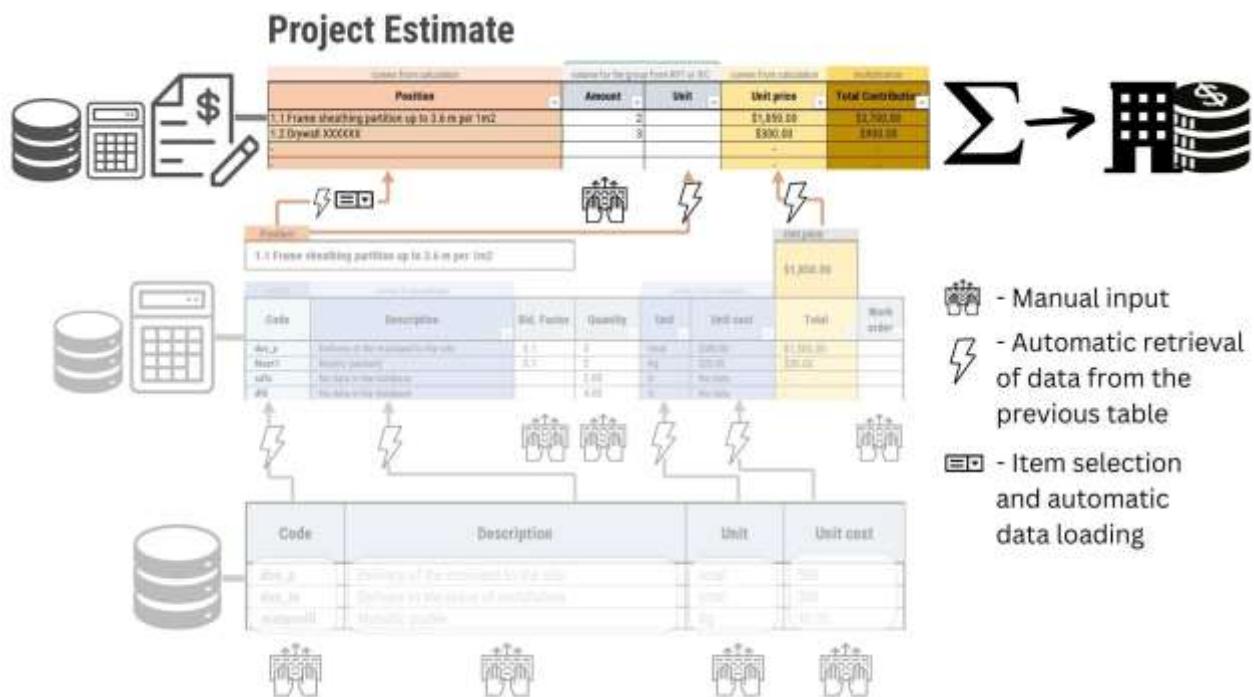
V důsledku toho může úroveň konkurence, poptávka na trhu, cílové marže a dokonce i vztahy s konkrétními zákazníky vést k výrazným odchylkám od standardizovaných norem. Tato praxe poskytuje flexibilitu trhu, ale také ztěžuje transparentní porovnávání nabídek různých dodavatelů a vnáší do stavebnictví v této fázi kalkulačního procesu prvek spekulativních cen.

Jakmile jsou připraveny kalkulační šablony pro jednotlivé činnosti a procesy - nebo častěji jednoduše zkopírovány ze standardních státních odhadů (obr. 5.1-7) s přidanými koeficienty, které odrážejí "zvláštnosti" konkrétní společnosti - je posledním krokem vynásobení nákladů každé položky odpovídajícím atributem rozsahu prací nebo procesů v novém projektu.

Při výpočtu celkových nákladů na nový stavební projekt je klíčovým krokem shrnutí nákladů na všechny položky kalkulace, které se vynásobí objemem těchto položek prací v projektu.

Pro vytvoření celkových nákladů projektu začneme v našem zjednodušeném příkladu výpočtem nákladů na vybudování jednoho metru čtverečního stěny a náklady na její výpočet (např. práce "1m² standardní instalace stěnových prvků") vynásobíme celkovým počtem metrů čtverečních stěn v projektu (např. atribut "Plocha" nebo "Množství" (obr. 5.1-8) entity typu "Stěnové prvky" z CAD projektu nebo výpočty mistra).

Podobně vypočítáme náklady na všechny prvky projektu (obr. 5.1-8): vezmeme náklady na jednotku práce a vynásobíme je objemem konkrétního prvku nebo skupiny prvků v projektu. Odhadce musí pouze zadat počet těchto prvků, činností nebo procesů v projektu na ve formě objemu nebo množství. To umožňuje automaticky vygenerovat kompletní odhad stavby.



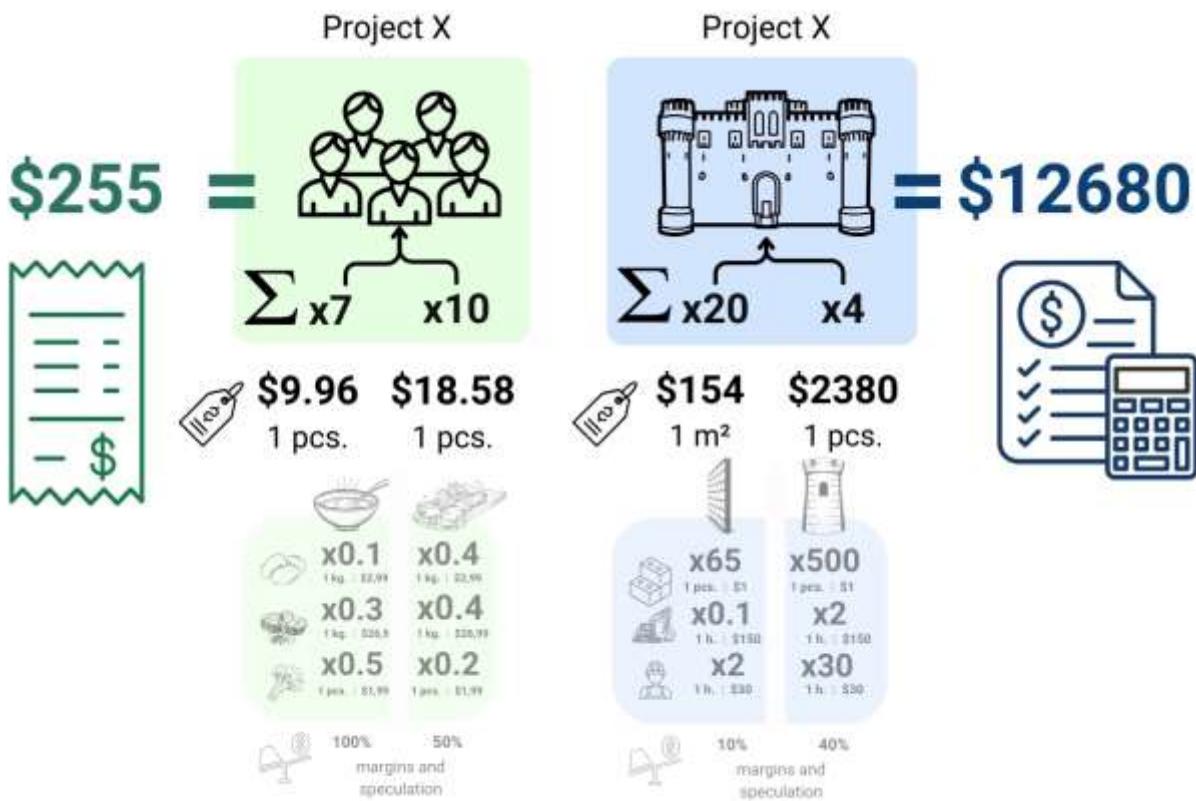
Obr. 5.1-8 Ve fázi tvorby odhadu zadáváme pouze rozsah prací.

Stejně jako v případě kalkulací i na této úrovni nahráváme automaticky hotové kalkulované položky (ze šablony kalkulace nebo nové zkopiované ze šablony a upravené), které s sebou automaticky nesou aktuální náklady na jednotku práce (které se automaticky aktualizují z databáze zdrojů (obr. 5.1-8 spodní tabulka)). V souladu s tím se v případě jakékoli změny údajů v databázi zdrojů nebo v tabulkách kalkulace nákladů - údaje v odhadu automaticky aktualizují pro aktuální den, aniž by bylo nutné měnit kalkulaci nákladů nebo samotný odhad.

V restauraci se konečné náklady na akci počítají podobným způsobem a rovnají se konečným nákladům na celou večeři, kdy se náklady na jednotlivé pokrmy vynásobené počtem hostů sečtou do celkové ceny šeku (obrázek 5.1-9). A stejně jako ve stavebnictví se recepty na vaření v restauraci nemusí měnit po celá desetiletí. Na rozdíl od cen, kde se náklady na ingredience mohou měnit každou hodinu.

Stejně jako majitel restaurace vynásobí náklady na každé jídlo počtem porcí a osob, aby určil celkové náklady na akci, manažer odhadující náklady sečte náklady na všechny součásti projektu a získá kompletní odhad stavby.

Pro každou činnost v projektu se tedy určí její konečný náklad (obr. 5.1-9), který se vynásobí objemem atributu entity odpovídající této činnosti - získá se náklad skupiny činností, z něhož se získá konečný náklad celého projektu.



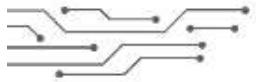
Obrázek 5.1-9 Konečný odhad se vypočítá tak, že se atribut nákladů na práci každého prvku sečte s jeho atributem rozsahu.

Celkové náklady projektu (obrázek 5.1-8) poskytují finanční obraz projektu, který umožňuje zákazníkům, investorům nebo finančním organizacím pochopit celkový rozpočet a finanční zdroje potřebné k realizaci projektu v daný den s ohledem na aktuální ceny.

A jestliže procesy sestavování zdrojových podkladů, kalkulací a odhadů (procesních receptů) jsou již vypracovány, poloautomatizovány a zdokonalovány desetitisíce let a evidovány na státní úrovni, pak automatické získávání kvalitativních informací o objemu a množství prvků pro poslední fázi konečného odhadu - dnes zůstává úzkým místem v procesech všech kalkulací nákladových a časových atributů projektu a obecně celkového rozpočtu projektu.

Tradiční metodou výpočtu objemů byly po tisíce let ruční metody měření objemu a množství pomocí plochých výkresů. S příchodem digitálního věku společnosti zjistily, že informace o objemu a množství lze nyní automaticky získávat z geometrických dat obsažených v modelech CAD, čímž došlo k revoluci v tisíciletých způsobech získávání kvantitativních údajů.

Moderní přístupy k odhadování a oceňování procesů zahrnují automatickou extrakci objemových a kvantitativních atributů z databází CAD, které lze nahrát a připojit k procesu kalkulace nákladů, aby bylo možné získat aktuální objemy projektových skupin v jakékoli fázi návrhu až po provoz.



KAPITOLA 5.2.

VÝPOČET MNOŽSTVÍ A AUTOMATICKÉ VYTVAŘENÍ ODHADŮ A HARMONOGRAMŮ

Přechod z 3D na 4D a 5D: využití objemových a kvantitativních parametrů

Když máte v ruce kalkulační tabulky s popsanými procesy prostřednictvím zdrojů (obr. 5.1-8), je dalším krokem automatické získání objemových nebo množstevních parametrů pro skupinu prvků, které jsou potřebné pro výpočty a pro konečný odhad.

Objemové charakteristiky prvků projektu - např. stěn nebo desek - lze automaticky získat z databáze CAD. Parametrické objekty vytvořené v programech CAD jsou pomocí geometrického jádra převedeny na číselné hodnoty délky, šířky, plochy, objemu a dalších parametrů. Procesu získávání objemů na základě 3D geometrie se budeme podrobněji věnovat v další, šesté části (obr. 6.3-3), věnované práci s CAD (BIM). Kromě objemů lze z databáze CAD-modelu získat také počet podobných prvků pomocí filtrování a seskupování objektů podle kategorií a vlastností. Tyto parametry, které umožňují seskupování, se stávají základem pro propojení prvků projektu prostřednictvím výpočtu zdrojů s kalkulacemi, konečnými odhady a rozpočtem celého projektu.

Datový model získaný z 3D (CAD) modelu je tak rozšířen o nové vrstvy parametrů, označované jako 4D a 5D. V nových vrstvách atributů entit, 4D (čas) a 5D (náklady), se jako zdroj hodnot atributů objemu entit používají 3D geometrická data.

- **4D** je vrstva informací o parametrech, která k 3D parametrym prvků přidává informace o době trvání stavebních operací. Tyto údaje jsou nezbytné pro plánování harmonogramů a řízení časového plánu projektu
- **5D** je další úrovní rozšíření datového modelu, ve kterém jsou prvky doplněny o nákladové charakteristiky. Geometrické informace se tak doplňují o finanční aspekt: náklady na materiál, práce a vybavení, což umožňuje výpočty rozpočtu, analýzy ziskovosti a řízení nákladů v průběhu výstavby.

Nákladové a 3D, 4D a 5D atributové údaje skupin subjektů projektu jsou popsány podobně jako kalkulace v modulárních ERP, PIMS -systémech (nebo nástrojích podobných Excelu) a slouží k automatickému plánování nákladů a rozpočtu jak jednotlivých skupin, tak i celého rozpočtu projektu.

5D atributy a získávání objemů atributů z CAD

Při přípravě konečného odhadu stavebního projektu, jehož přípravě jsme se věnovali v předchozích kapitolách (obrázek 5.1-8), se objemové atributy pro jednotlivé kategorie prvků projektu shromažďují buď ručně, nebo se získávají ze specifikací objemových atributů poskytovaných softwarem CAD.

Tradiční manuální metoda výpočtu množství zahrnuje analýzu výkresů, které se po tisíce let předkládají jako čáry na papíře a posledních 30 let v digitálních formátech, jako je PDF (PLT) nebo DWG. Vycházejí z odborných zkušeností a měří množství potřebných prací a materiálů, často pomocí pravítka a úhloměru. Tato metoda vyžaduje značné úsilí a čas, stejně jako zvláštní pozornost věnovanou detailům.

Stanovení atributů rozsahu tímto způsobem může trvat od několika dnů až po několik měsíců v závislosti na rozsahu projektu. Navíc vzhledem k tomu, že veškerá měření a výpočty se provádějí ručně, existuje riziko lidské chyby, která může vést k nepřesným údajům, což má následně vliv na chybu v odhadu času a nákladů na projekt, za které ponese odpovědnost celá společnost.

Moderní metody založené na využití databází CAD výrazně zjednoduší výpočet objemů. V modelech CAD již geometrie prvků obsahuje atributy objemu, které lze automaticky vypočítat (prostřednictvím jádra geometrie (obr. 6.3-3)) a prezentovat nebo exportovat v tabulkové podobě.

V takovém případě oddělení odhadu požádá projektanta CAD o údaje o množství a objemových charakteristikách prvků projektu. Tato data se exportují ve formě tabulek nebo se přímo integrují do kalkulačních databází - ať už jde o Excel, ERP nebo PMIS - systémy. Tento proces často nezačíná formální žádostí, ale krátkým dialogem mezi klientem (iniciátorem) a architektem odhadcem ze stavební nebo projekční firmy. Níže je uveden zjednodušený příklad, který ukazuje, jak z každodenní komunikace vzniká strukturovaná tabulka pro automatické výpočty (QTO):

- ⌚ Zákazník - "Chci k budově přistavět další patro ve stejné konfiguraci jako první patro"
- ⌚ Architekt (CAD) - "Přidání třetího patra, konfigurace je stejná jako ve druhém patře". A po této zprávě odešle novou verzi projektu v CADu odhadci.
- ⌚ Odhadce automaticky provede seskupení a výpočet (ERP, PMIS, Excel) - "Provedu projekt přes tabulku v Excelu s pravidly QTO (ERP, PMIS), získám objemy podle kategorií pro novou podlahu a vygeneruji odhad"

Výsledkem je transformace textového dialogu do struktury tabulky s pravidly pro seskupování:

Prvek	Kategorie	Podlaha
Překrytí	OST_Floors	3

Prvek	Kategorie	Podlaha
Sloupec	OST_StructuralColumns	3
schodiště	OST_Stairs	3

Po procesu automatického seskupení modelu CAD od projektanta podle pravidel QTO odhadce a automatického vynásobení objemů výpočtem zdrojů (obr. 5.1-8) získáme následující výsledky, které jsou odeslány zákazníkovi:

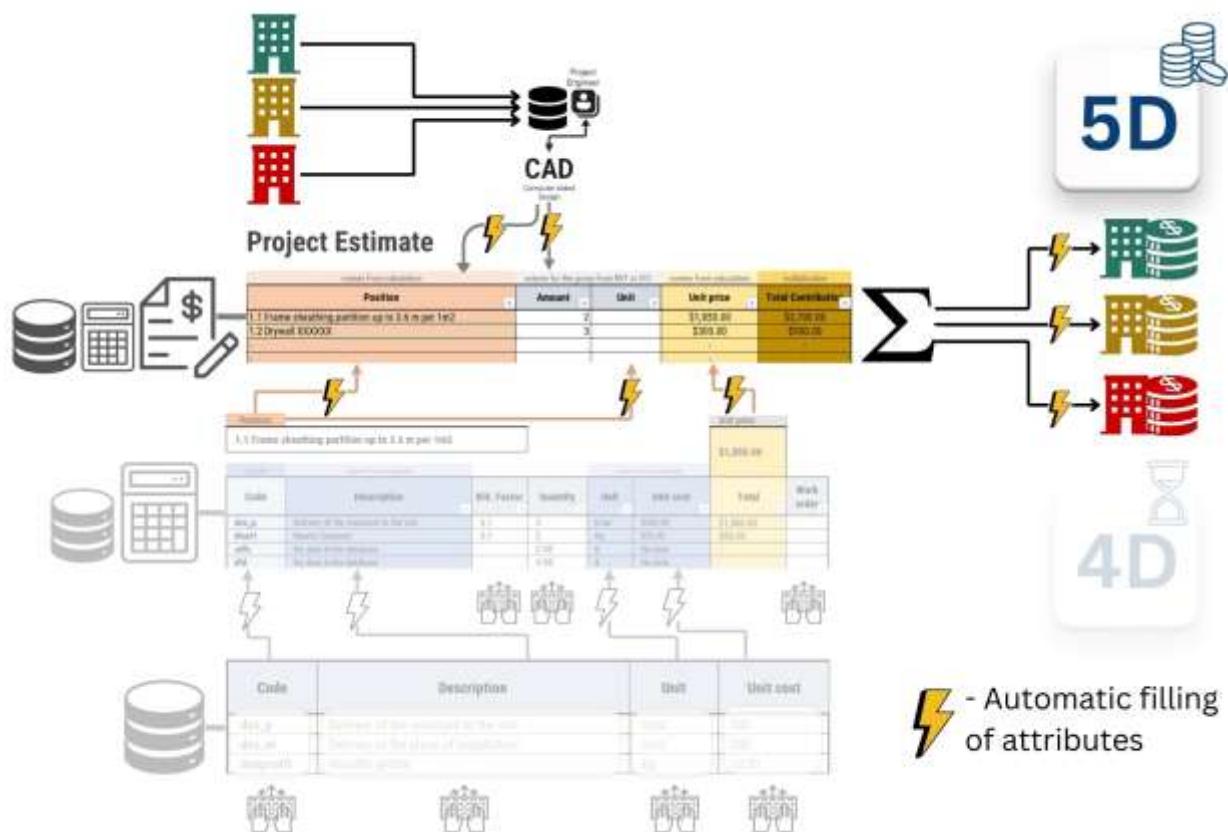
Prvek	Svazek	Podlaha	Cena za jednotku.	Celkové náklady
Překrytí	420 m ²	3	150 €/m ²	63 000 €
Sloupec	4 ks.	3	2450 €/kus.	9 800 €
schodiště	2 ks.	3	4 300 €/kus.	8 600 €
CELKEM:	-	-	-	81 400 €

- 💡 Zákazník - "Děkuji, je toho docela hodně, potřebujeme zkrátit několik místností". A cyklus se opakuje mnohokrát.

Tento scénář se může opakovat mnohokrát, zejména ve fázi schvalování, kdy zákazník očekává okamžitou zpětnou vazbu. V praxi se však takové procesy mohou protáhnout na dny nebo dokonce týdny. Dnes by díky zavedení automatického seskupování a pravidel pro výpočet měly být činnosti, které dříve zabraly značný čas, hotové během několika minut. Automatizované získávání veličin prostřednictvím pravidel pro seskupování nejen urychluje výpočty a odhady, ale minimalizací lidského faktoru snižuje pravděpodobnost chyb a poskytuje transparentní a přesné posouzení nákladů na projekt.

Pokud byly požadavky odhadního oddělení původně zohledněny při vytváření 3D modelu v systému CAD (což je v praxi stále vzácné) a názvy, identifikátory skupin prvků a jejich klasifikační atributy jsou nastaveny ve formě parametrů, které se shodují se strukturami odhadních skupin a tříd, pak lze objemové atributy automaticky přenést do odhadních systémů bez dalších transformací.

Automatická extrakce objemových atributů z CAD ve formě specifikačních tabulek umožňuje rychle získat aktuální údaje o nákladech jednotlivých prací a projektu jako celku (obr. 5.2-1). Pouhou aktualizací souboru CAD s objemy projektu v procesu výpočtu nebo v kalkulačním systému může firma rychle přepočítat odhad s ohledem na nejnovější změny, což zajistí vysokou přesnost a konzistenci všech následných výpočtů.



Obr. 5.2-1 Atributy objemu z tabulek CAD nebo databází jsou automaticky zaneseny do odhadu, což umožňuje okamžitý výpočet celkových nákladů projektu.

S rostoucí složitostí investičních projektů se výpočet celého rozpočtu a analýza celkových nákladů projektů podle takového scénáře (obrázek 5.2-1) - stává klíčovým nástrojem pro informované rozhodování.

Podle studie společnosti Accenture Creating More Value through Capital Projects (2024) [20] přední společnosti aktivně integrují datovou analytiku do digitálních iniciativ a využívají historické informace k předvídaní a optimalizaci výsledků. Výzkum ukazuje, že stále více vlastníků a provozovatelů používá analýzu velkých dat k předvídání tržních trendů a posuzování komerční životoschopnosti před zahájením projektování. Toho je dosaženo analýzou datových skladů ze stávajícího portfolia projektů. Kromě toho 79% vlastníků-provozovatelů zavádí "robustní" prediktivní analytiku pro hodnocení výkonnosti projektů a podporu provozního rozhodování v reálném čase.

Moderní efektivní řízení stavebních projektů je neoddělitelně spjato se zpracováním a analýzou velkého množství informací ve všech fázích projektování a procesů, které projektování předcházejí. Využití datových skladů, výpočtů zdrojů, prediktivních modelů a strojového učení umožňuje nejen minimalizovat rizika při výpočtech, ale také přijímat strategická rozhodnutí o financování projektů v raných fázích projektování. O datových skladech a prediktivních modelech, které doplní výpočty, budeme více hovořit v deváté části knihy.

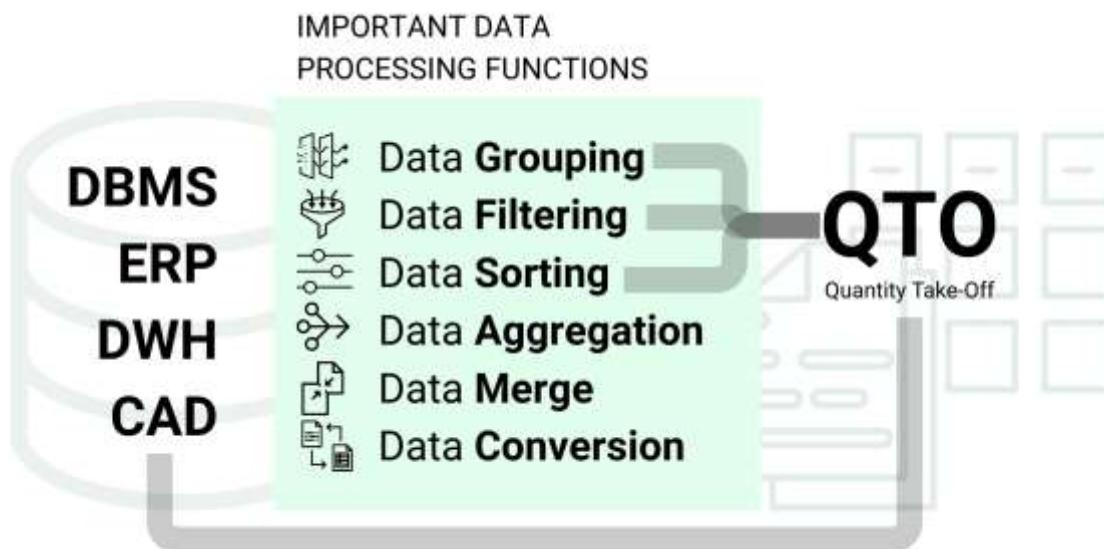
Automatické získávání objemových parametrů prvků z projektů CAD, které jsou nezbytné pro přípravu

odhadů, se provádí pomocí nástrojů pro seskupování QTO (Quantity Take-Off). Nástroje QTO pracují tak, že seskupují všechny objekty projektu podle speciálních identifikátorů prvků nebo parametrů atributů prvků s využitím specifikací a tabulek vytvořených v databázi CAD.

QTO Quantity Take-Off: seskupení dat projektu podle atributů.

QTO (Quantity Take-Off) ve stavebnictví je proces stanovení množství prvků potřebných k realizaci projektu. V praxi je QTO často polomanuální proces zahrnující sběr dat z různých zdrojů: dokumentů PDF, výkresů DWG a digitálních modelů CAD.

Při práci s daty získanými z databází CAD je proces QTO realizován jako sled operací filtrování, třídění, seskupování a agregace. Prvky modelu se vybírají podle parametrů tříd, kategorií a typů, načež se jejich kvantitativní atributy - například objem, plocha, délka nebo množství - shrnují podle logiky výpočtu (obr. 5.2-2).



Obr. 5.2-2 Seskupování a filtrování dat jsou nejoblíbenější funkce používané v databázích a datových skladech.

Proces QTO (filtrování a seskupování) umožňuje systematizovat data, vytvářet specifikace a připravovat vstupní informace pro výpočet odhadů, nákupů a harmonogramů prací. Základem QTO je klasifikace prvků podle typu měřených atributů. Pro každý prvek nebo skupinu prvků se vybere odpovídající kvantitativní parametr měření. Např:

- **Atribut délky** (obrubník - v metrech)
- **Přívlastek plochy** (sádrokartonářské práce - v metrech čtverečních)
- **Příznak objemu** (betonářské práce - v metrech krychlových)
- **Příznak množství** (okna - na kus)

Kromě objemových charakteristik generovaných matematicky na základě geometrie se ve výpočtech

po seskupení QTO často používají faktory překročení (obr. 5.2-12, např. 1,1 pro zohlednění 10% pro logistiku a instalaci) - korekční hodnoty, které zohledňují ztráty, instalaci, skladování nebo dopravu. To umožňuje přesněji předvídat skutečnou spotřebu materiálů a vyhnout se nedostatku i nadměrným zásobám na staveništi.

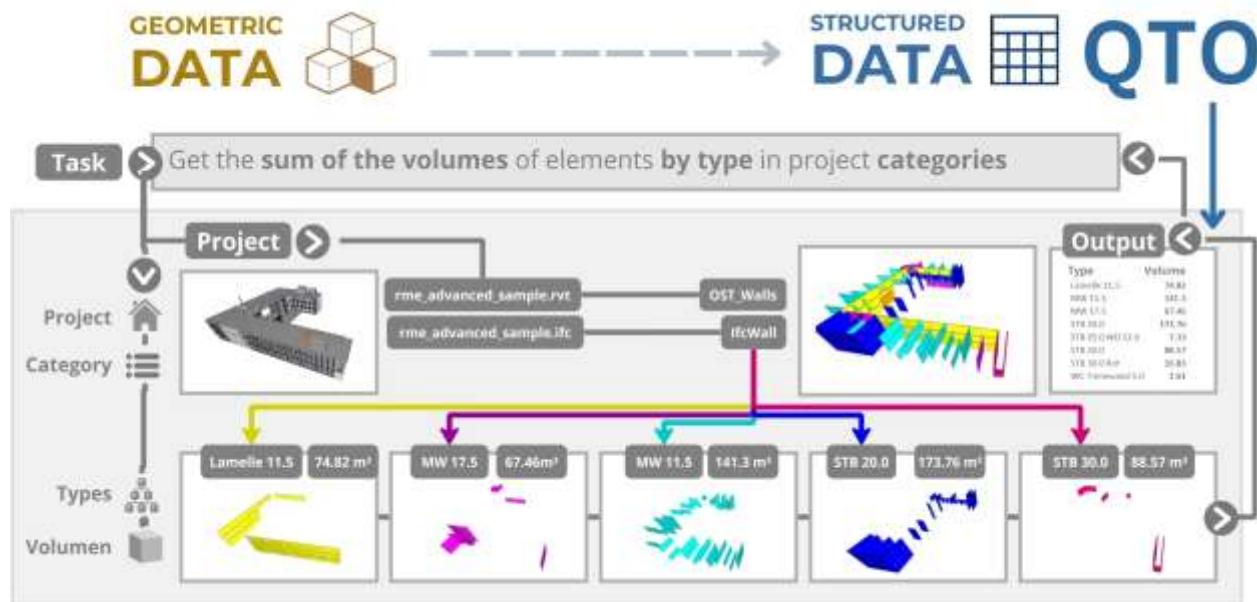
Automatizovaný proces odběru množství (QTO) je nezbytný pro vytvoření přesných výpočtů a odhadů na, omezení lidských chyb v procesech specifikace objemu a prevenci nadměrného nebo nedostatečného objednání materiálů.

Jako příklad postupu QTO uveďme běžný případ, kdy je třeba z databáze CAD zobrazit tabulku se specifikací objemů podle typů prvků pro určitou kategorii, třídu prvků. Seskupme všechny prvky projektu podle typu z kategorie stěny projektu CAD a shrňme atributy objemů pro každý typ, abychom výsledek prezentovali jako tabulku objemů QTO (obr. 5.2-3).

V příkladu typického projektu CAD (obr. 5.2-3) jsou všechny prvky kategorie stěn v databázi CAD seskupeny podle typu stěny, např. "Lamelle 11.5", "MW 11.5" a "STB 20.0", a mají přesně definované objemové atributy reprezentované v metrických kostkách.

Cílem manažera, který je na rozhraní mezi projektanty a specialisty na výpočty, je získat automatizovanou tabulku objemů podle typu prvku ve vybrané kategorii. A to nejen pro konkrétní projekt, ale i v univerzální podobě použitelné pro další projekty s podobnou strukturou modelu. Díky tomu je tento přístup škálovatelný a umožňuje opakované použití dat bez duplicitního úsilí.

Pryč jsou doby, kdy se zkušení projektanti a odhadci vyzbrojovali pravítkem a pečlivě měřili každou čáru na papíře nebo v PDF -plánech - tradice, která se za poslední tisíciletí nezměnila. S rozvojem 3D modelování, kde je nyní geometrie každého prvku přímo spojena s automaticky vypočítanými objemovými atributy, se proces určování objemů a QTO veličin zautomatizoval.



Obrázek 5.2-3 Získání atributů objemu a množství QTO z projektu zahrnuje seskupování a filtrování prvků projektu.

V našem příkladu je úkolem "vybrat kategorie stěn v projektu, seskupit všechny prvky podle typu a prezentovat informace o objemových atributech ve strukturovaném, tabulkovém formátu" tak, aby tuto tabulku mohly využít desítky dalších specialistů pro výpočty kalkulací, logistiky, harmonogramů prací a dalších obchodních případů (obr. 6.1-3).

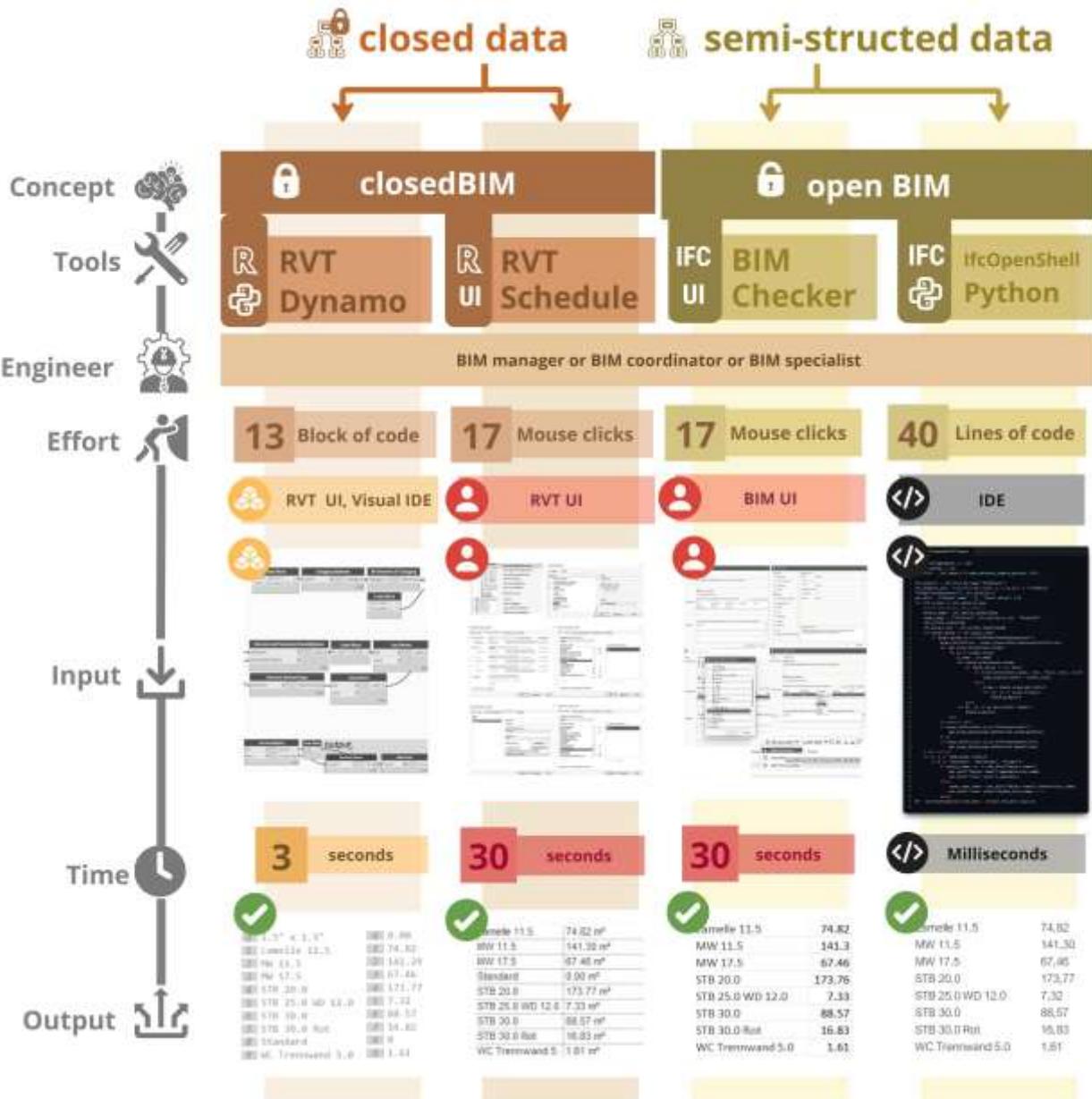
Vzhledem k uzavřenému charakteru dat CAD nemůže dnes každý odborník využít přímý přístup do databáze CAD (důvody a řešení problému přístupu jsou podrobně popsány v šesté části knihy). Proto se mnozí musí obracet na specializované nástroje BIM založené na konceptu otevřeného BIM a uzavřeného BIM [63]. Při práci se specializovanými nástroji BIM nebo přímo v prostředí softwaru CAD může být tabulka s výsledky QTO (Quantity Take-Off) generována různými způsoby - podle toho, zda je použito ruční rozhraní, nebo softwarová automatizace.

Například při použití uživatelského rozhraní softwaru CAD (BIM) stačí provést přibližně 17 akcí (kliknutí na tlačítko), abyste získali hotovou tabulku objemů (obr. 5.2-4). Uživatel však musí dobré znát strukturu modelu a funkce softwaru CAD (BIM).

Pokud se automatizace použije prostřednictvím programového kódu nebo zásuvných modulů a nástrojů API v programech CAD, sníží se počet ručních kroků k získání objemových tabulek, ale v závislosti na použité knihovně nebo nástroji bude třeba napsat 40 až 150 řádků kódu:

- **IfcOpSh (otevřený BIM)** nebo **Dynamo IronPython (uzavřený BIM)** - umožňují získat tabulku QTO z formátu CAD nebo CAD programu na pouhých ~40 řádcích kódu.
- **IFC_js (otevřený BIM)** - vyžaduje přibližně 150 řádků kódu pro extrakci objemných atributů z modelu IFC.
- **Nástroje rozhraní CAD (BIM)** - umožňují dosáhnout stejného výsledku ručně, na 17 kliknutí

myší.



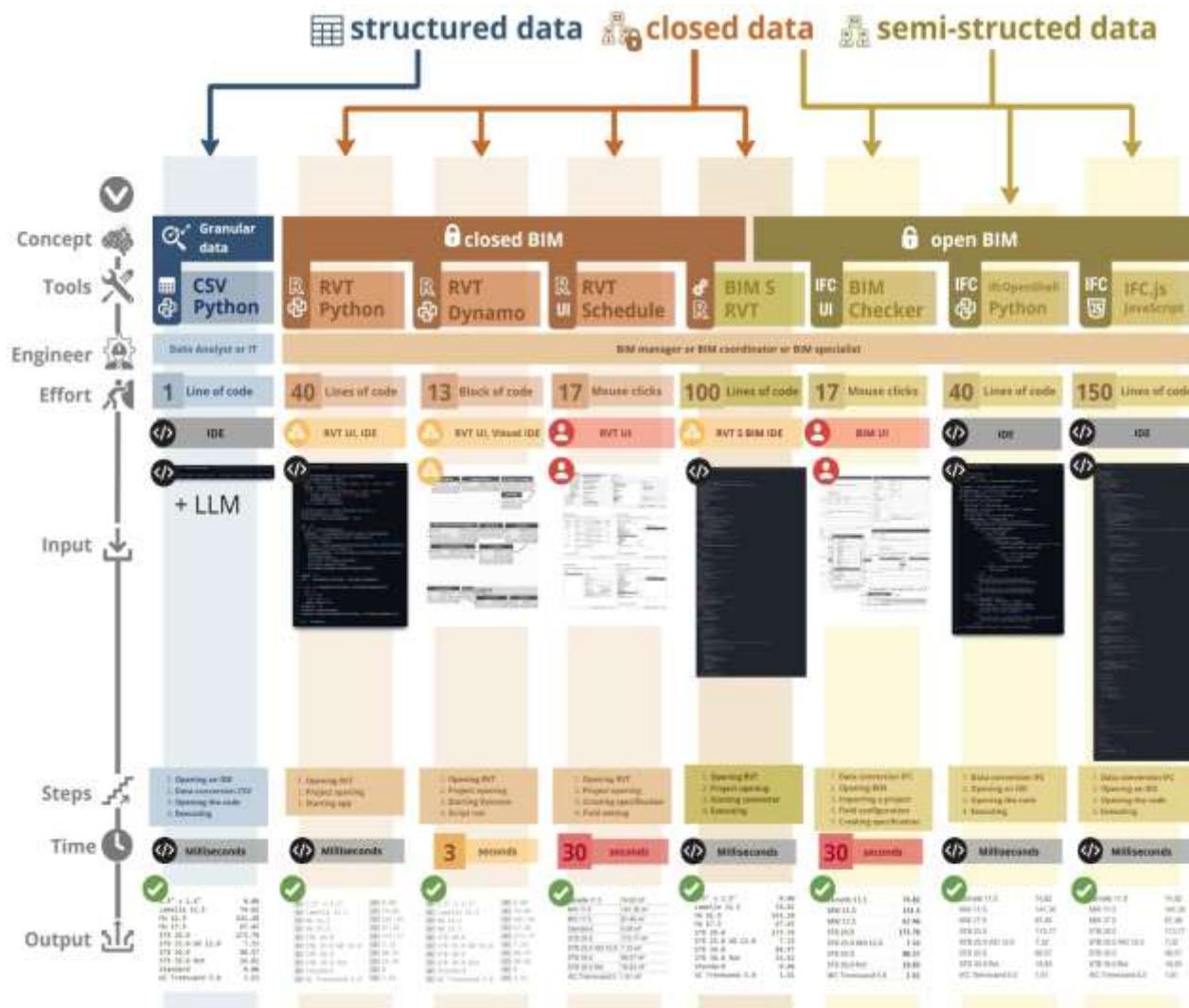
Obrázek 5.2-4 CAD (BIM) projektanti a manažeři, používají 40 až 150 řádků kódu nebo desítky stisků kláves k vytvoření QTO tabulek.

Výsledek je stejný - strukturovaná tabulka s atributy svazku pro skupinu prvků. Jediným rozdílem jsou náklady na práci a úroveň technických znalostí vyžadovaných uživatelem (obr. 5.2-4). Moderní nástroje oproti ručnímu sběru objemů výrazně urychlují proces QTO a snižují pravděpodobnost chyb. Umožňují získávat údaje přímo z modelu projektu, čímž odpadá nutnost ručního přepočítávání objemů z výkresů, jak se to dělalo v minulosti.

Bez ohledu na použitou metodu - zda otevřený BIM nebo uzavřený BIM - je možné získat identický QTO

- tabulku s objemy prvků projektu (obr. 5.2-4). Při práci s daty projektu v CAD - (BIM-) koncepcí jsou však uživatelé závislí na specializovaných nástrojích a API poskytovaných dodavatelem (obr. 3.2-13). To vytváří další vrstvy závislosti a vyžaduje naučit se jedinečná datová schémata a zároveň omezuje přímý přístup k datům.

Vzhledem k uzavřenému charakteru dat CAD komplikuje pořizování QTO tabulek a dalších parametrů automatizaci výpočtů a integraci s externími systémy. Použitím nástrojů pro přímý přístup k databázím a převedením CAD -dat projektu pomocí nástrojů reverzního inženýrství do otevřeného formátu strukturovaného datového rámce (obr. 4.1-13) lze získat identickou QTO tabulkou pomocí jediného řádku kódu (obr. 5.2-5 - varianta s granulovanými daty).



Obrázek 5.2-5 Různé nástroje poskytují stejné výsledky v podobě atributových tabulek subjektů projektu, ale s různými náklady na práci.

Při použití otevřených strukturovaných dat z projektů CAD, jak je uvedeno v kapitole "Převod dat CAD (BIM) do strukturované podoby", se výrazně zjednoduší proces seskupování, QTO.

Přístupy založené na využití otevřených strukturovaných dat nebo přímém přístupu k databázím

modelů CAD jsou oproštěny od marketingových omezení spojených se zkratkou BIM. Opírají se o osvědčené nástroje dlouhodobě používané v jiných odvětvích (obr. 7.3-10. Proces ETL).

Podle studie společnosti McKinsey "Open Data: Uvolněte inovace a produktivitu pomocí streamovaných informací". [102] provedené v roce 2013 by využívání otevřených dat mohlo vytvořit příležitosti k úsporám ve výši 30 až 50 miliard dolarů ročně v oblasti projektování, inženýrství, zadávání veřejných zakázek a výstavby energetických zařízení. To znamená 15procentní úsporu investičních nákladů na výstavbu.

Práce s otevřenými strukturovanými (granulovanými) daty zjednoduší vyhledávání a zpracování informací, snižuje závislost na specializovaných platformách BIM a otevírá cestu k automatizaci bez nutnosti používat proprietární systémy nebo parametrické a složité datové modely z formátů CAD.

Automatizace QTO pomocí LLM a strukturovaných dat

Převod nestrukturovaných dat do strukturované podoby výrazně zvyšuje efektivitu různých procesů: zjednoduší vyhledávání a zpracování dat (obr. 4.1-1, obr. 4.1-2) a urychlí proces validace tím, že požadavky jsou jasné a přehledné, jak jsme již uvedli v předchozích kapitolách. Podobně převod dat CAD (BIM) do strukturované otevřené formy (obr. 4.1-12, obr. 4.1-13) usnadňuje proces seskupování atributů a proces QTO.

Tabulka atributů QTO má strukturovanou podobu, takže při použití strukturovaných dat CAD pracujeme s jediným datovým modelem (obrázek 5.2-5), čímž odpadá nutnost převádět a konvertovat datové modely projektů a pravidla seskupování na společného jmenovatele. To nám umožňuje seskupovat data podle jednoho nebo více atributů pomocí jediného řádku kódu. Naproti tomu v otevřeném BIM a uzavřeném BIM, kde jsou data uložena v polostrukturovaných, parametrických nebo uzavřených formátech, vyžaduje zpracování desítky nebo dokonce stovky řádků kódu a použití rozhraní API pro interakci s geometrií a atributovými informacemi.

- ❷ Příklad seskupení strukturovaného projektu QTO podle jednoho atributu. Textový dotaz v libovolné chatovací místnosti LLM (ChatGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN nebo jiné):

Mám CAD -project jako DataFrame - prosím filtroujte data projektu tak, abyste získali položky s parametrem "Type" obsahujícím pouze hodnotu "Type 1" ↵

- Odpověď LLM bude s velkou pravděpodobností v podobě kódu Pythonu pomocí Pandas :



Obr. 5.2-6 Jeden řádek kódu napsaný pomocí LLM , umožňuje seskupit celý projekt CAD podle atributu "Type" a získat požadovanou skupinu prvků.

Díky jednoduché struktuře dvourozměrného DataFrame nemusíme vysvětlovat schéma a datový model LLM, což zkracuje výkladové kroky a urychluje tvorbu finálních řešení. Dříve vyžadovalo psaní i jednoduchého kódu učení programovacích jazyků, ale nyní nám moderní jazykové modely (LLM) umožňují automaticky převádět logiku procesu do kódu při práci se strukturovanými daty pomocí textových dotazů.

Automatizace a jazykové modely LLM mohou zcela eliminovat potřebu profesionálů pracujících s daty CAD (BIM) seskupovat a zpracovávat data, aby se museli učit programovací jazyky nebo nástroje BIM, protože poskytují možnost řešit problémy pomocí textových dotazů.

Stejný dotaz - seskupení všech prvků projektu z kategorie "stěny" a výpočet objemů pro každý typ (obr. 5.2-5) - který v prostředí CAD (BIM) vyžaduje 17 kliknutí v rozhraní nebo napsání 40 řádků kódu, v otevřených nástrojích pro zpracování dat (např. SQL nebo Pandas) vypadá jako jednoduchý a intuitivní dotaz:

- Pomocí jediného řádku v programu Pandas:

```
df[df['Category'].isin(['OST_Walls'])].groupby('Type')['Volume'].sum()
```

Dekódování kódu: Vezměte z df (DataFrame) prvky, jejichž atribut-sloupec "Category", má hodnoty "OST_Walls", seskupte všechny získané prvky podle atributu-sloupce "Type" a sečtěte pro získanou skupinu prvků atribut "Volume".

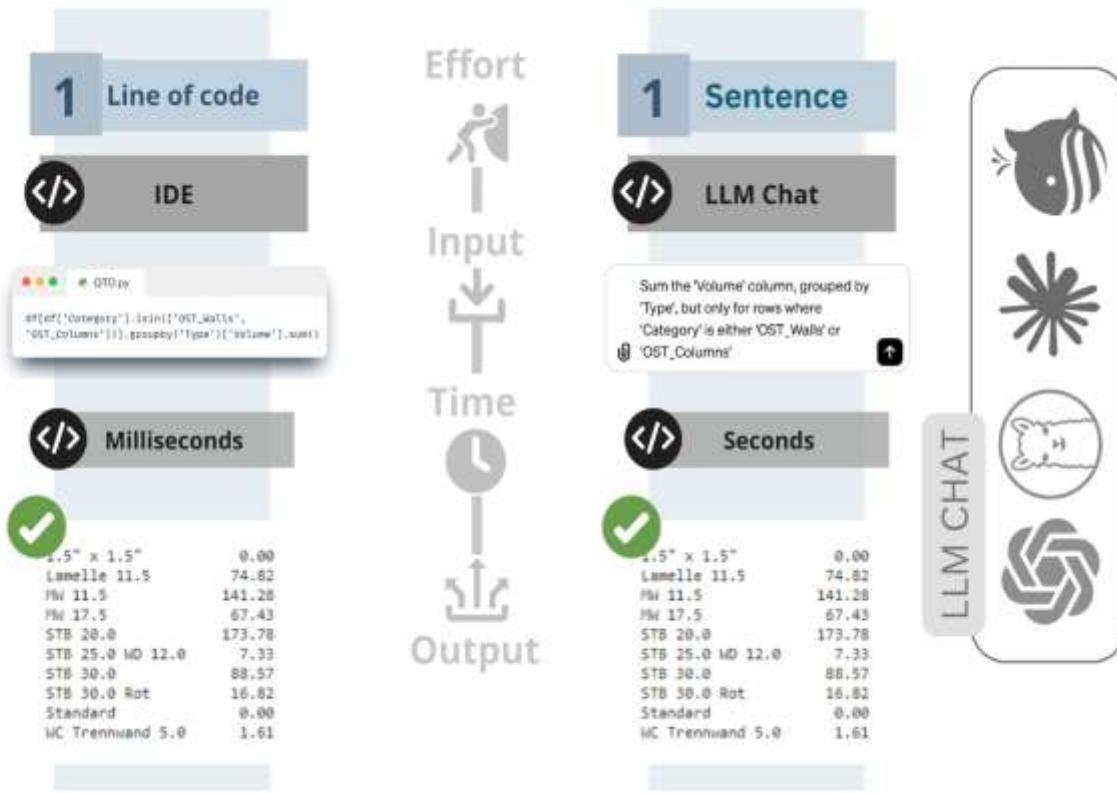
- Seskupení strukturovaného projektu získaného z CADu pomocí SQL:

```
SELECT Typ, SUMA(Objem) AS Celkový objem
FROM prvky
WHERE Category = 'OST_Walls'
GROUP BY Typ;
```

- Pomocí LLM můžeme zapsat požadavek na seskupení do databáze projektu jako jednoduchý textový odkaz - výzvu (obr. 5.2-7):

Pro datový rámec projektu seskupte položky podle parametru "Typ", ale pouze pro položky s parametrem "Kategorie" rovným "OST_Stěny" nebo "OST_Sloupce", a pro výslednou sečtěte parametr sloupce "Objem". ↵

PYTHON PANDAS

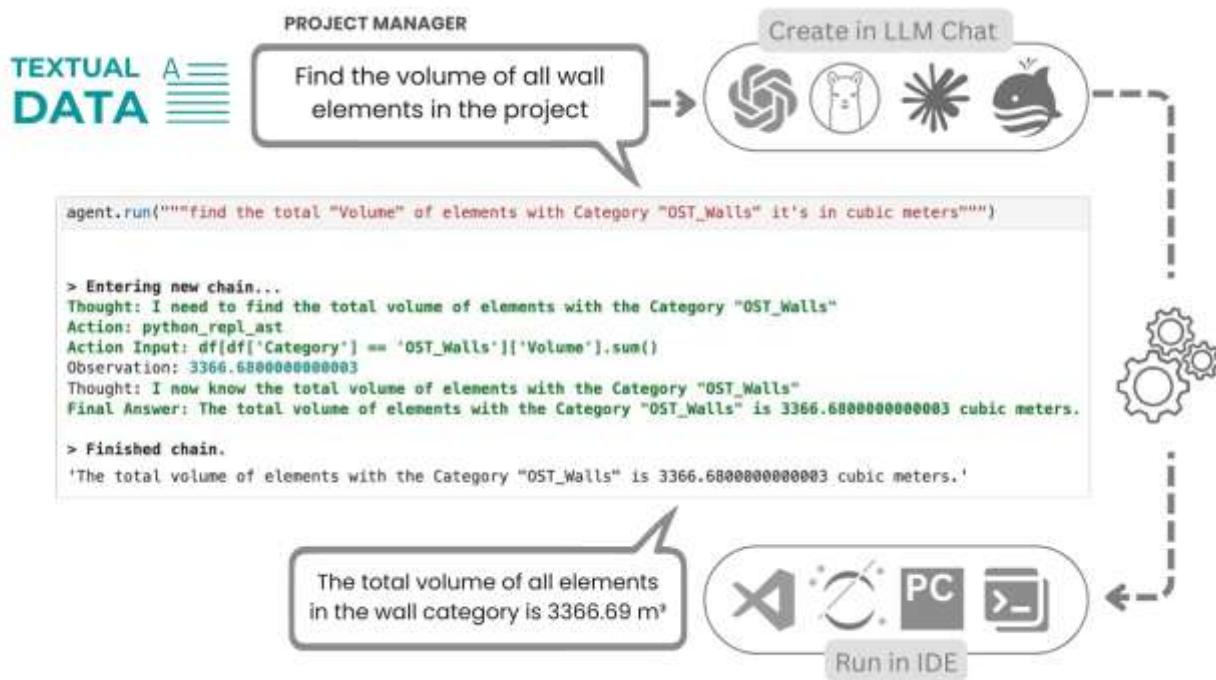


Obrázek 5.2-7 Pomocí SQL, Pandas a LLM je nyní možné automatizovat zpracování dat pomocí několika řádků kódu a textových dotazů.

Získávání QTO z dat CAD pomocí nástrojů LLM (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok), dramaticky mění tradiční metody získávání atributových informací, kvantitativních a objemových dat pro jednotlivé objekty a skupiny objektů.

Nyní mohou i projektoví manažeři, specialisté na kalkulaci nákladů nebo logistiku, kteří nemají hluboké znalosti projektování a nemají specializovaný software CAD - (BIM-) dodavatelé, kteří mají přístup k databázi CAD, získat celkový objem prvků kategorie stěn nebo jiných objektů během několika sekund, a to pouhým napsáním nebo nadiktováním dotazu.

V textových dotazech (obr. 5.2-8) agent LLM modelu zpracovává požadavek uživatele aplikaci určité funkce na jeden nebo více parametrů - sloupců tabulky. Výsledkem je, že uživatel při komunikaci s LLM obdrží buď nový sloupec-parametr s novými hodnotami, nebo jednu konkrétní hodnotu po seskupení



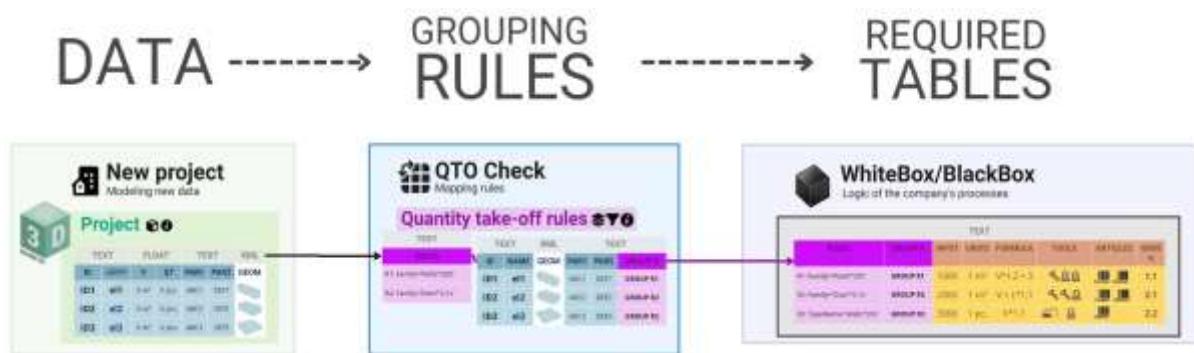
Obrázek 5.2-8 Model LLM, který pracuje se strukturovanými daty, z kontextu textového dotazu pochopí, na jaké seskupení a atributy se uživatel ptá.

Pokud je třeba získat veličiny pouze pro jednu skupinu prvků, stačí provést jednoduchý dotaz QTO (obrázek 5.2-7) na data modelu CAD. Při výpočtu rozpočtu nebo odhadu pro celý projekt sestávající z mnoha skupin prvků je však často nutné získat množství pro všechny typy prvků (třídy0, přičemž každá kategorie prvků se zpracovává samostatně - se seskupením podle příslušných atributů).

V praxi odhadců a posuzovatelů se pro různé typy objektů používají jednotlivá pravidla pro seskupování a výpočet. Například okna se obvykle seskupují podle podlaží nebo zón (parametr seskupení - atribut Úroveň, Místo) a stěny podle materiálu nebo typu konstrukce (parametr Materiál, Typ). Aby bylo možné proces seskupování automatizovat, jsou tato pravidla předem popsána na adrese ve formě tabulek seskupovacích pravidel. Tyto tabulky fungují jako konfigurační šablony, které definují, které atributy mají být použity při výpočtech pro jednotlivé skupiny prvků v projektu.

QTO výpočet celého projektu pomocí skupinových pravidel z tabulky Excelu.

V reálných stavebních projektech je často nutné provádět agregaci podle několika atributů současně v rámci jedné skupiny prvků. Například při práci s kategorií "Okna" (kde atribut Kategorie obsahuje hodnoty jako OST_Okna nebo IfcOkna) lze prvky seskupovat nejen podle typu - například podle hodnoty v poli Název typu nebo Typ - ale také podle dalších charakteristik, jako je například úroveň tepelné vodivosti uvedená v příslušném atributu. Toto vícerozměrné seskupování umožňuje získat přesnější výsledky pro konkrétní skupinu. Podobně při výpočtu kategorií stěn nebo podlah lze jako filtry nebo kritéria pro seskupení použít libovolné kombinace atributů - například materiál, úroveň, podlaha, požární odolnost a další parametry (obrázek 5.2-9).



Obrázek 5.2-9 Pro každou skupinu nebo kategorii subjektů v projektu existuje jiný vzorec pro seskupení, který se skládá z jednoho nebo více kritérií.

Proces definování takových pravidel seskupování je podobný procesu tvorby požadavků na data popsanému v kapitole "Tvorba požadavků a kontrola kvality dat" (obrázek 4.4-5), kde jsme se podrobně zabývali prací s datovými modely. Taková pravidla seskupování a výpočtu zajišťují přesnost a relevantnost výsledků pro automatický výpočet celkových atributů množství nebo objemu kategorie entit s přihlédnutím ke všem nezbytným podmínkám, které je třeba při výpočtech a kalkulacích zohlednit.

- ❷ Následující ukázka kódu filtrouje tabulku projektů tak, aby výsledná datová sada obsahovala pouze entity, u nichž sloupec atributů "Category" obsahuje hodnoty "OST_Windows" nebo "IfcWindows" a zároveň sloupec atributů "Type" obsahuje hodnotu "Type 1":

Mám projekt DataFrame - filtrovat data tak, aby v datasetu zůstaly pouze položky, které mají atribut "Category" obsahující hodnoty "OST_Windows" nebo "IfcWindows" a zároveň atribut Type obsahuje hodnotu "Type 1".

💡 Odpověď společnosti LLM :

```
group.py
1 df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['Type'].str.contains("Type 1"))]
```

Obrázek 5.2-10 Jediný řádek kódu, podobný vzorci aplikace Excel , umožňuje seskupit všechny entity projektu podle několika atributů.

Výsledný kód (obr. 5.2-10) po překladu dat CAD ve strukturovaných otevřených formátech (obr. 4.1-13) lze spustit v některém z populárních IDE (integrovaných vývojových prostředí), která jsme zmínili výše, v offline režimu: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse s pluginem PyDev, Thonny, Wing IDE, IntelliJ IDEA s pluginem Python, JupyterLab nebo populární online nástroje: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

💡 Pro získání entit projektu ve formuláři QTO DataFrame v kategorii "Windows" pouze s konkrétní hodnotou tepelné vodivosti můžeme použít následující dotaz do LLM:

Mám projekt DataFrame - filtruji data tak, aby v datasetu zůstaly pouze záznamy s "Category" obsahující hodnoty "OST_Windows" nebo "IfcWindows" a zároveň aby sloupec ThermalConductivity měl hodnotu 0,. ↴

💡 Odpověď společnosti LLM :

```
group.py
1 df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['ThermalConductivity'] == 0.5)]
```

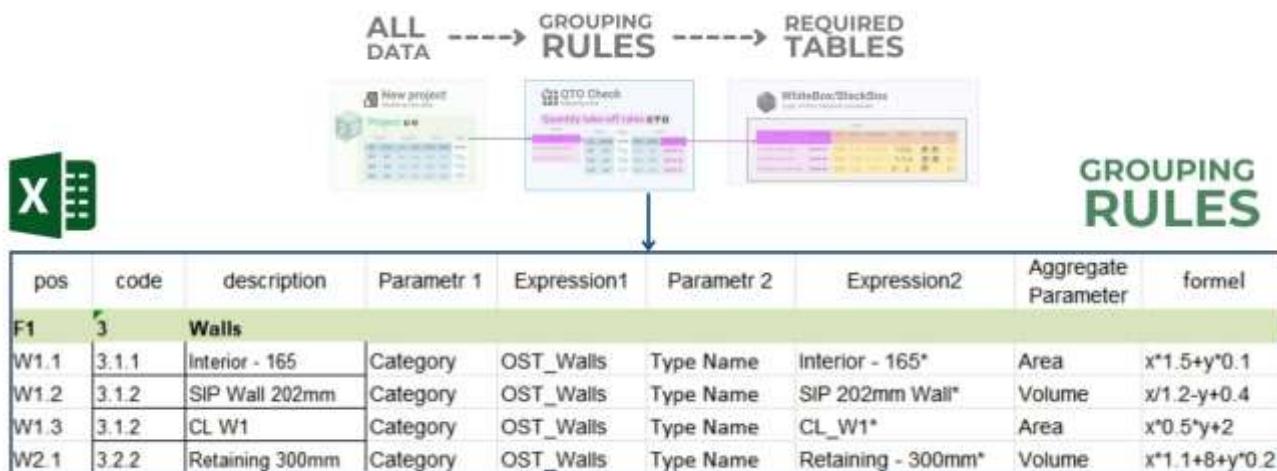
Obrázek 5.2-11 Extrémně jednoduchý dotazovací jazyk Pandas Python umožňuje spouštět QTO pro libovolný počet projektů současně.

V odpovědi obdržené od LLM (obr. 5.2-11) je použita logická podmínka "&", která kombinuje dvě kritéria: hodnotu tepelné vodivosti a příslušnost k jedné ze dvou kategorií. Metoda "isin" kontroluje, zda je hodnota atributu-sloupce "Category" obsažena v poskytnutém seznamu.

V projektech s velkým počtem skupin prvků s různou logikou seskupování - pro každou kategorii entit projektu (např.: okna, dveře, desky) by měla být stanovena individuální pravidla seskupování, která mohou zahrnovat další koeficienty nebo vzorce pro výpočet celkových atributů. Tyto vzorce (obrázek 5.2-12 atributový "formel", např. x-hodnota množství a y-objem skupiny) a koeficienty zohledňují jedinečné vlastnosti každé skupiny, např:

- % přírůstek objemu materiálu pro zohlednění překročení.
- pevné dodatečné množství materiálu
- úpravy související s možnými riziky a chybami ve výpočtech v podobě vzorců.

Jakmile jsou pravidla filtrování a seskupování formulována ve formě vzorců parametrů pro každou kategorii položek, lze je uložit jako tabulku po řádcích - například ve formátu Excel (obrázek 5.2-12). Uložením těchto pravidel ve strukturované podobě lze proces výběru, filtrování a seskupování dat projektu plně automatizovat. Namísto ručního psaní mnoha samostatných dotazů systém jednoduše načte tabulku parametrů a aplikuje příslušná pravidla na model (celkový datový rámec projektu (obr. 4.1-13)), čímž vytvoří konečné tabulky QTO pro každou kategorii prvků projektu.



Obrázek 5.2-12 Tabulka pro seskupování atributů QTO stanovuje pravidla pro seskupování prvků projektu a poskytuje přesný celkový počet a rozsah pro každou kategorii.

Shromážděná pravidla vám umožní seskupit celý projekt a provést všechny potřebné výpočty, včetně úprav atributů objemu. Výsledkem je přivedení objemů do "skutečného objemu", který se používá pro výpočty a kalkulace, a ne těch, které byly původně ve fázi návrhu v modelu CAD.

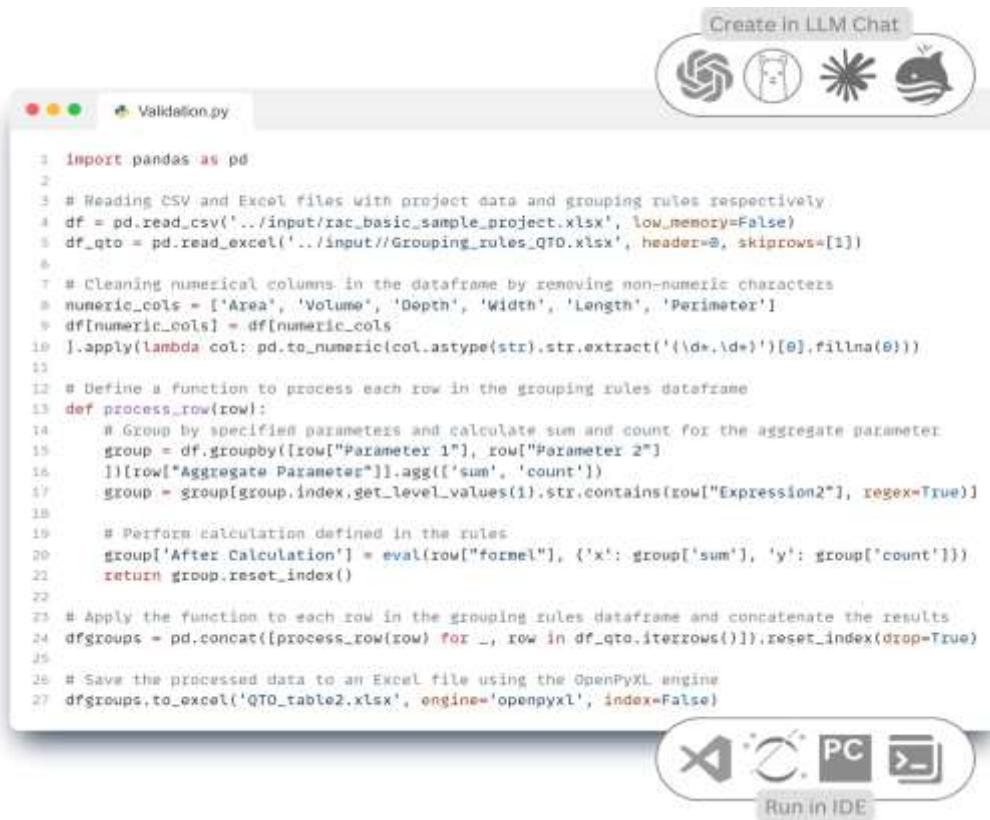
V procesu automatického vytváření objemových tabulek QTO pro celý projekt musí aplikace projít všechny kategorie tabulky pravidel seskupování, vzít atributy seskupování, seskupit podle nich všechny prvky projektu a agregovat atribut objemu pro tuto skupinu a navíc jej vynásobit zpřesňujícím faktorem nebo koeficientem.

Požádejme LLM, aby nám napsal kód pro takové řešení, kde by kód můžel načíst dvě tabulky - tabulku pravidel pro seskupování (obrázek 5.2-12) a tabulku dat ze samotného projektu (obrázek 4.1-13), a poté aplikovat pravidla pro seskupování, seskupit položky podle daných pravidel, vypočítat agregované hodnoty a výsledky uložit do nového souboru Excelu.

✉ Pošlete textovou žádost na chat LLM:

Potřebuji kód, který načte data projektu ze souboru 'basic_sample_project.xlsx' a poté pravidla ze souboru 'Grouping_rules_QTO.xlsx', podle kterých se seskupí všechna data podle 'Parametr 1' a 'Parametr 2', agreguje se 'Agregat Parameter', filtruje se podle 'Expression2', provedou se výpočty z 'Formel1' a uloží se tabulka QTO do 'QTQ_table2.xlsx' ↴

❑ Odpověď společnosti LLM :



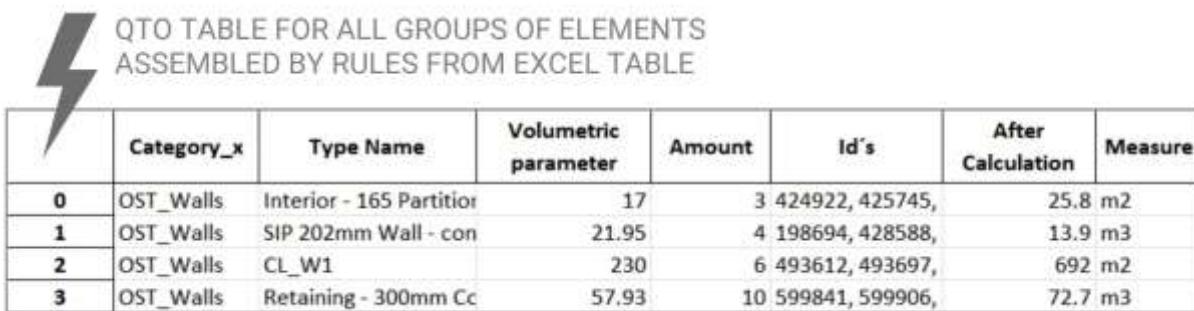
```

1 import pandas as pd
2
3 # Reading CSV and Excel files with project data and grouping rules respectively
4 df = pd.read_csv('../input/rac_basic_sample_project.xlsx', low_memory=False)
5 df_qto = pd.read_excel('../input//Grouping_rules_QTO.xlsx', header=0, skiprows=[1])
6
7 # Cleaning numerical columns in the dataframe by removing non-numeric characters
8 numeric_cols = ['Area', 'Volume', 'Depth', 'Width', 'Length', 'Perimeter']
9 df[numeric_cols] = df[numeric_cols]
10 df.apply(lambda col: pd.to_numeric(col.astype(str).str.extract('(\d+\.\d+)'))[0], fillna(0))
11
12 # Define a function to process each row in the grouping rules dataframe
13 def process_row(row):
14     # Group by specific parameters and calculate sum and count for the aggregate parameter
15     group = df.groupby([row["Parameter 1"], row["Parameter 2"]])
16     l1[row["Aggregate Parameter"]].agg(['sum', 'count'])
17     group = group[group.index.get_level_values(1).str.contains(row["Expression2"])]
18
19     # Perform calculation defined in the rules
20     group['After_Calculation'] = eval(row["formula"], {'x': group['sum'], 'y': group['count']})
21     return group.reset_index()
22
23 # Apply the function to each row in the grouping rules dataframe and concatenate the results
24 dfgroups = pd.concat([process_row(row) for _, row in df_qto.iterrows()]).reset_index(drop=True)
25
26 # Save the processed data to an Excel file using the OpenPyXL engine
27 dfgroups.to_excel('QTO_table2.xlsx', engine='openpyxl', index=False)

```

Obrázek 5.2-13 Kód jazyka Python Pandas zpracovává data návrhu z projektu CAD a seskupuje je podle pravidel definovaných v aplikaci Excel .

Konečným výsledkem provedení kódu (obr. 5.2-13) bude tabulka skupiny entit, která bude obsahovat nejen zobecněné atributy objemu z původního modelu CAD- (BIM-), ale také nový atribut reálného objemu, který zohledňuje všechny požadavky pro správné vytvoření výpočtů a odhadů (příklad obr. 5.2-14).

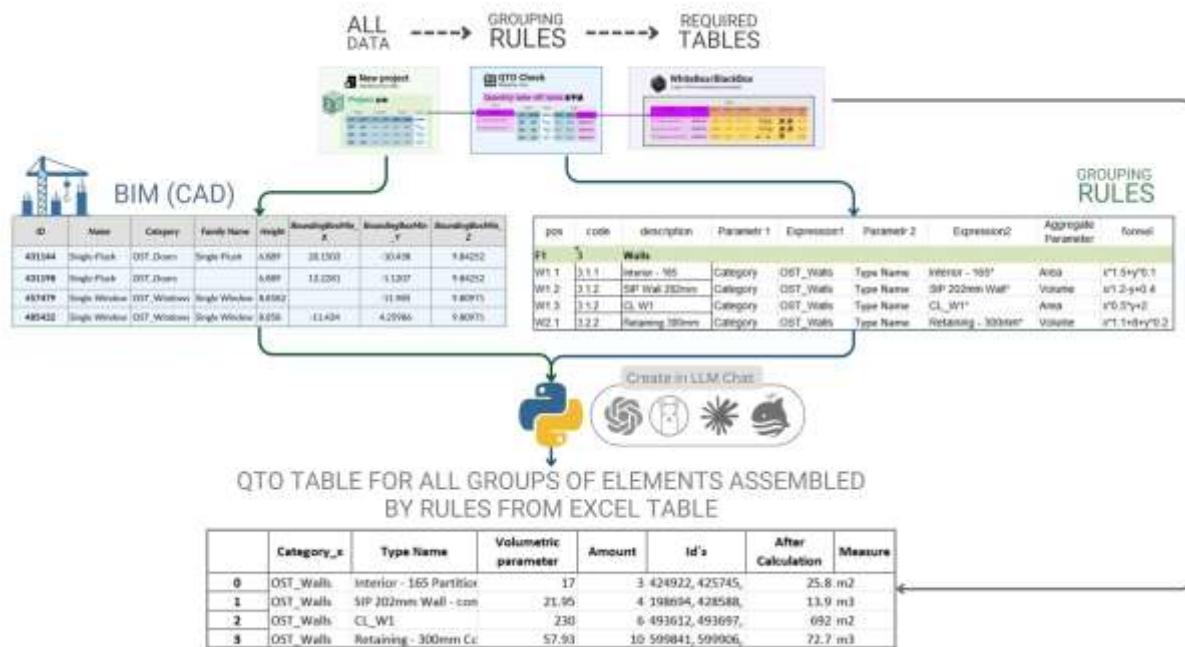


	Category_x	Type Name	Volumetric parameter	Amount	Id's	After Calculation	Measure
0	OST_Walls	Interior - 165 Partition	17	3 424922, 425745,		25.8 m2	
1	OST_Walls	SIP 202mm Wall - con	21.95	4 198694, 428588,		13.9 m3	
2	OST_Walls	CL_W1	230	6 493612, 493697,		692 m2	
3	OST_Walls	Retaining - 300mm Cc	57.93	10 599841, 599906,		72.7 m3	

Obrázek 5.2-14 Atribut "Po výpočtu" se přidá do souhrnné tabulky po provedení kódu, který automaticky vypočítá skutečný objem.

Výsledný kód (obr. 5.2-13) lze spustit v některém z populárních IDE (o kterých jsme se zmínili výše) a

aplikovat jej na libovolný počet stávajících nebo nových příchozích projektů (RVT, IFC, DWG, NWS, DGN atd.), ať už se jedná o několik projektů nebo třeba stovky projektů v různých formátech ve strukturované podobě (obr. 5.2-15).



Obrázek 5.2-15 Proces automatického seskupování stavebních dat propojuje data BIM (CAD) s tabulkami QTO prostřednictvím pravidel z tabulky Excel.

Přizpůsobený a parametrizovaný proces sběru objemových dat (obrázek 5.2-15) umožňuje plně automatizovaný sběr dat o kvantitativních atributech a objemech prvků projektu pro další práci s nimi, včetně odhadu nákladů, logistiky, harmonogramů prací a výpočtu uhlíkové stopy a dalších analytických úloh.

Po osvojení nástrojů, které nám umožňují snadno organizovat a seskupovat skupiny prvků projektu podle určitých atributů, jsme nyní připraveni integrovat seskupené a filtrované projekty s různými podnikovými výpočty a obchodními scénárii.



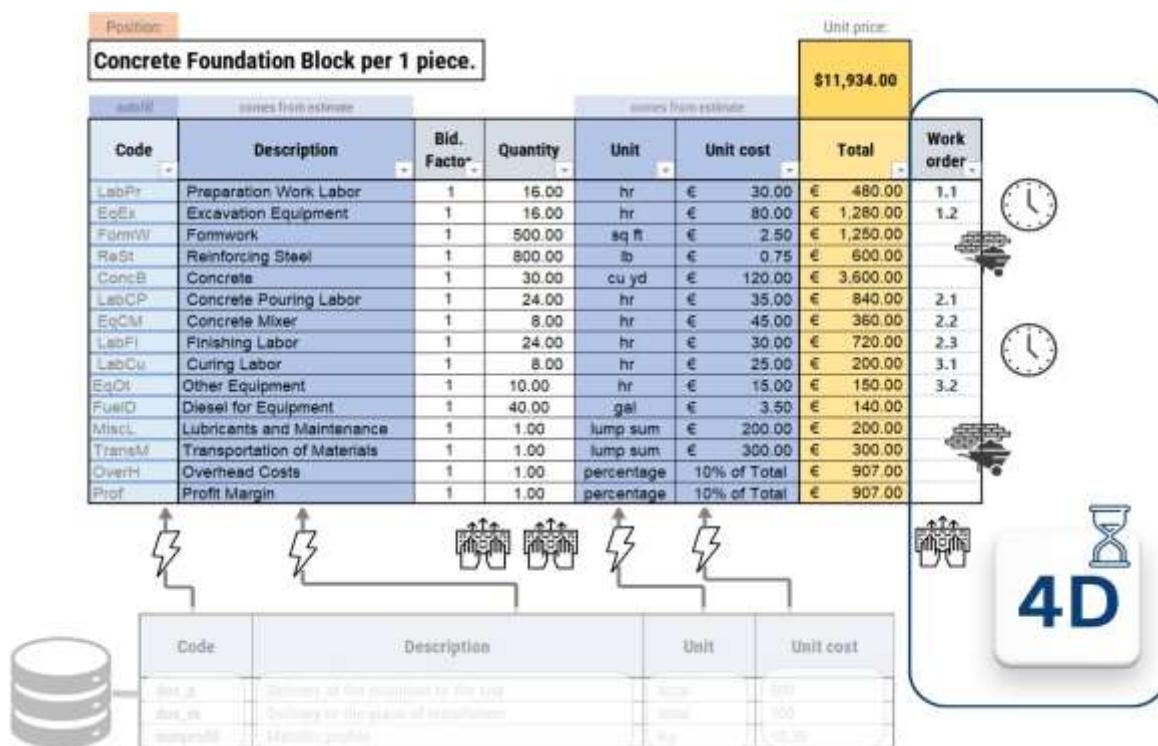
KAPITOLA 5.3.

4D, 6D -8D A VÝPOČET EMISÍ OXIDU UHLIČITÉHO CO₂.

4D model: integrace času do odhadů stavby

Kromě kalkulace nákladů je jednou z klíčových aplikací projektových dat ve stavebnictví stanovení časových parametrů - jak pro jednotlivé stavební operace, tak pro celý projekt. Jako podklad pro automatizovaný výpočet časových parametrů a tvorbu harmonogramu provádění prací se často používá metoda odhadu podle zdrojů a s ní spojená kalkulační databáze, o níž bylo podrobně pojednáno v předchozí kapitole "Kalkulace a odhady pro stavební projekty".

Přístup založený na zdrojích zohledňuje nejen materiálové náklady, ale také časové zdroje. Při kalkulaci nákladů lze každému procesu přiřadit atribut pracovního příkazu (obrázek 5.3-1 - Parametr pracovního příkazu) a množství času a nákladů spojených s provedením tohoto procesu. Tyto parametry jsou důležité zejména pro popis operací, které nemají pevnou tržní cenu a nelze je přímo koupit - například použití stavebních zařízení, zaměstnání dělníků nebo logistické procesy (které se obvykle vyjadřují běžně v hodinách). V takových případech náklady neurčuje oddělení veřejných zakázek, ale přímo realizační firma na základě interních norem nebo výrobních sazeb (obrázek 5.3-1).



Obr. 5.3-1 Výpočty práce při metodě odhadu podle zdrojů zahrnují časové rozvržení pracovních hodin.

Výpočty na úrovni kalkulace nákladů tak zahrnují nejen náklady na pohonné hmoty a materiál (pořizovací náklady), ale také čas strávený řidiči, techniky a pomocnými pracovníky na místě. V

uvedeném příkladu (obr. 5.3-1) je v tabulce nákladů uvedena kalkulace nákladů na instalaci základového bloku, která zahrnuje jednotlivé složky práce, jako je příprava, instalace rámu a betonáž, a také potřebný materiál a práci. Jednotlivé operace, jako jsou přípravné práce, však nemusí mít žádné materiálové náklady, ale mohou obsahovat značné dočasné mzdové náklady vyjádřené v člověkohodinách.

Pro plánování pořadí prací (pro harmonogram prací) na staveništi se do výpočtové tabulky ručně přidá atribut "Pracovní zakázka" (obr. 5.3-1). Tento atribut je uveden v dalším sloupci pouze u položek, jejichž měrná jednotka je vyjádřena v čase (hodina, den). Tento atribut doplňuje kód práce, popis, množství, měrnou jednotku (parametr "Jednotka") a náklady. Číselná posloupnost (parametr "Pořadí prací") činností umožňuje stanovit pořadí, v jakém se úkoly na staveništi provádějí, a použít jej pro plánování.

Harmonogram výstavby a jeho automatizace na základě údajů o nákladech

Harmonogram výstavby představuje vizuální znázornění plánu prací a procesů, které budou prováděny v rámci realizace projektu. Vytváří se na základě podrobných výpočtů zdrojů (obrázek 5.3-1), kde jsou jednotlivé úkoly-práce naplánovány kromě nákladů na zdroje také podle času a pořadí.

Na rozdíl od zprůměrovaných přístupů, kdy se časové odhady zakládají na typických hodinách pro instalaci materiálů nebo zařízení, se při metodě založené na zdrojích plánuje na základě skutečných údajů v kalkulaci nákladů. Každá položka odhadu týkající se mzdových nákladů vychází z použitého kalendáře, který zohledňuje skutečné podmínky využití zdrojů během pracovního období. Úprava produktivních hodin prostřednictvím koeficientů na úrovni kalkulace nákladů (obr. 5.3-1 parametr "Bid faktor"), umožňuje zohlednit rozdíly v produktivitě a sezonní zvláštnosti ovlivňující časový průběh prací.

Abychom určili data zahájení a ukončení procesu pro harmonogram výstavby v Ganttově diagramu, vezmeme hodnoty atributu množství času pro každou položku z kalkulace nákladů na základové bloky a vynásobíme je počtem bloků (v tomto případě počtem betonových základových bloků). Tímto výpočtem získáme dobu trvání jednotlivých úkolů. Poté tyto doby trvání vyneseme na časovou osu, počínaje datem zahájení projektu, a vytvoříme graf, jehož výsledkem je vizuální znázornění, které ukazuje, kdy by měl každý úkol začít a skončit. Parametr "pořadí prací" u procesů nám navíc umožňuje pochopit, zda pracovní proces probíhá paralelně ("pořadí prací" např. 1.1-1.1) nebo sekvenčně (1.1-1.2).

Ganttův diagram je grafický nástroj pro plánování a řízení projektů, který znázorňuje úkoly jako vodorovné sloupce na časové ose. Každý sloupec představuje dobu trvání úkolu, jeho začátek a konec.

Harmonogram neboli Ganttův diagram pomáhá projektovým manažerům a pracovníkům jasně pochopit, kdy a v jakém pořadí mají být jednotlivé fáze výstavby prováděny, což zajišťuje efektivní využití zdrojů a dodržení termínů.

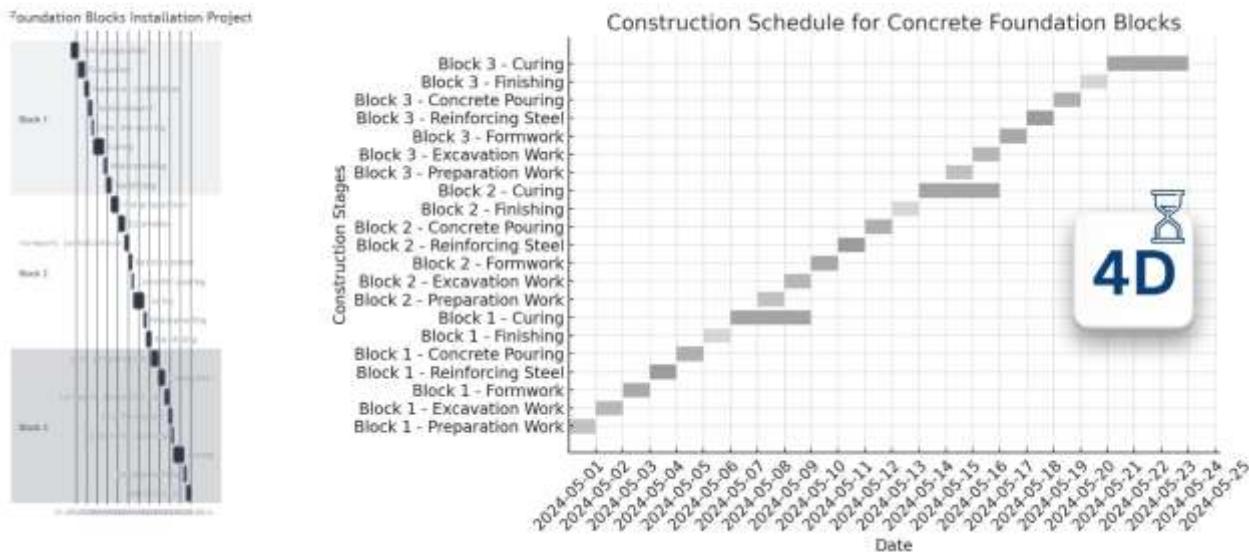
Uveděme časový plán pro instalaci tří betonových základových bloků s využitím výpočtů z výše uvedené tabulky. Na základě tabulky nákladů (obr. 5.3-1) z výše uvedeného příkladu požádejte LLM, aby sestavil harmonogram instalace 3 prvků základových bloků, např. na první květen 2024.

Pro odeslání kalkulace nákladů do LLM, můžeme nahrát kalkulační tabulku ve formátu XLSX nebo jednoduše vložit snímek obrazovky ve formátu JPEG kalkulace nákladů přímo do chatu LLM (obr. 5.3-1). LLM samostatně vyhledá knihovnu pro vizualizaci obrázku tabulky a vynásobením časových atributů činností z tabulky jejich objemem doplní všechny údaje do grafu.

- ✉ Pošlete textovou žádost na adresu LLM:

Vytvořte harmonogram prací ve formě Ganttova diagramu s použitím hodnot z tabulky časových vztahů (obr. 5.3-1 ve formátu JPEG v příloze) pro instalaci 3 základových bloků. Práce pro každý blok se provádějí postupně. Začátek prací se stanoví od 1. 5. 2024. ↴

- ✉ Odpověď společnosti LLM



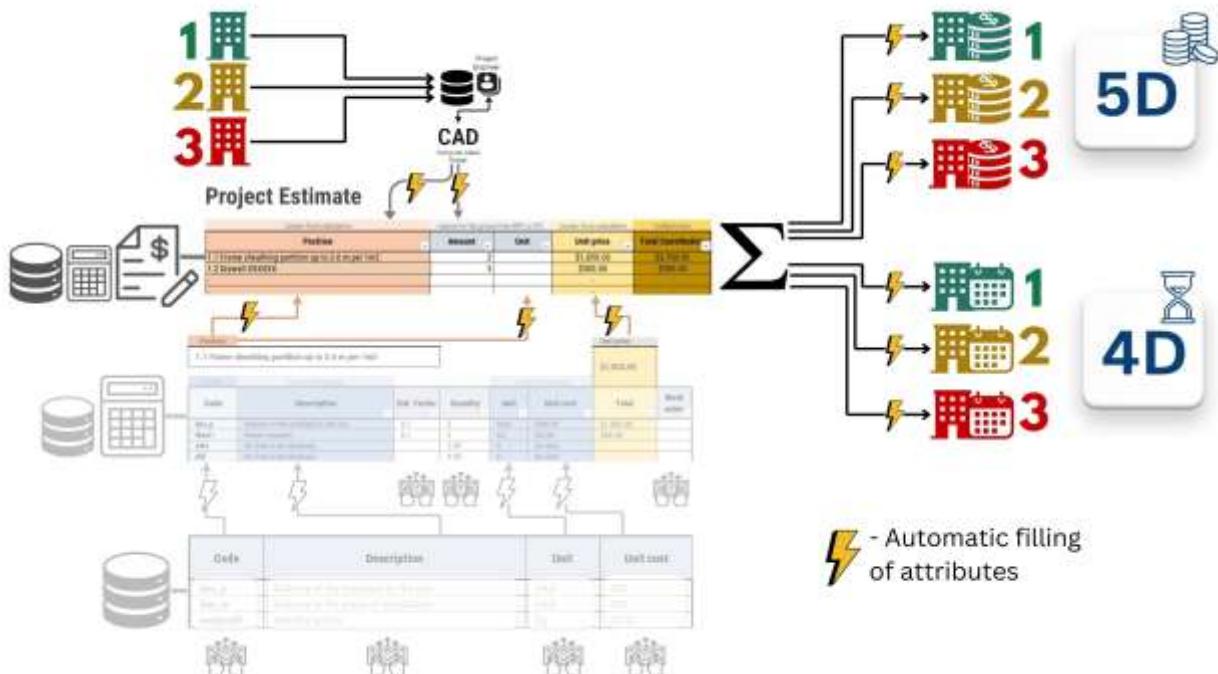
Obr. 5.3-2 Automaticky generovaný několika LLM Ganttuův diagram zobrazuje fáze výstavby tří betonových bloků podle podmínek z výzvy.

Výsledný graf (obr. 5.3-2) je časový diagram, ve kterém každý vodorovný sloupec odpovídá určité fázi prací na základovém bloku a zobrazuje sled operací (parametr "pořadí prací"), jako je příprava, výkop, montáž bednění, armování, betonáž a dokončovací práce, tj. ty procesy, které mají ve výpočtech vyplněné časové parametry a pořadí.

Takový harmonogram (obr. 5.3-2) nebude v úvahu omezení týkající se pracovních dnů, směn nebo norem pracovní doby, ale je určen pouze pro koncepční vizualizaci procesu. Přesný harmonogram, který bude odrážet souběžnost prací lze doplnit vhodnými podněty nebo dodatečnými pokyny v rámci chatu.

Pomocí jednotné kalkulace nákladů (obr. 5.3-1) je možné díky objemovým atributům z 3D -geometrie automaticky odhadnout jak náklady projektu prostřednictvím automatizovaných odhadů, tak zároveň vypočítat časové charakteristiky skupin ve formě tabulek nebo grafů pro různé varianty projektu (obr.

5.3-3).



Obr. 5.3-3 Automatický výpočet, umožňuje okamžité a automatické předvídání nákladů a času pro různé varianty projektu.

Moderní modulární ERP -systémy (obr. 5.4-4), které načítají data z modelů CAD, používají podobné automatizované metody výpočtu času, které výrazně zkracují rozhodovací proces. To umožňuje okamžitě a přesně plánovat harmonogramy prací a vypočítat celkový čas potřebný k dokončení všech úkolů při realizaci projektu s ohledem na reálné ceny.

Rozšířené atributové vrstvy 6D -8D: od energetické účinnosti po zajištění bezpečnosti

6D, 7D a 8D jsou rozšířené úrovně informačního modelování, z nichž každá přispívá dalšími vrstvami atributů do komplexního informačního modelu projektu, jehož základem jsou atributy 3D -modelu s jejich počtem a rozsahem. Každá další vrstva zavádí specifické parametry, které jsou potřebné pro další seskupování nebo další identifikaci v jiných systémech, jako jsou například systémy správy nemovitostí (PMS), počítačem podporovaná správa zařízení (CAFM), řízení stavebních projektů (CPM) a systémy řízení bezpečnosti (SMS).



Obrázek 5.3-4 Atributy 6D, 7D a 8D v datovém informačním modelu rozšiřují zohlednění různých aspektů projektu, od energetické účinnosti až po bezpečnost.

- **6D** jsou kromě databáze projektů (nebo datového rámce (obrázek 4.1-13)) s geometrickými a objemovými atributy prvků přidány informace (sloupce atributů) o environmentální udržitelnosti. Patří sem informace týkající se energetické účinnosti, uhlíkové stopy, recyklovatelnosti materiálů a používání technologií šetrných k životnímu prostředí. Tyto údaje umožňují posoudit dopad projektu na životní prostředí, optimalizovat rozhodnutí o projektu a dosáhnout cílů udržitelného rozvoje (ESG).
- **7D** atributy doplňují atributy potřebné pro řízení údržby budov. Jedná se o údaje o plánech údržby, životních cyklech komponent, technické dokumentaci a historii oprav. Tento soubor informací zajišťuje, že model lze integrovat se systémy údržby (CAFM, AMS), umožňuje efektivní plánování údržby, výměnu zařízení a poskytuje podporu po celou dobu životního cyklu objektu.
- **8D** další atributová vrstva, - zahrnuje informace týkající se bezpečnosti - jak ve fázi výstavby, tak během následného provozu. Model zahrnuje opatření pro bezpečnost personálu, pokyny pro případ nouze, evakuační systémy a požadavky na požární ochranu. Začlenění těchto údajů do digitálního modelu pomáhá předem zohlednit rizika a vypracovat architektonická, technická a organizační řešení, která zohledňují požadavky na bezpečnost a ochranu zdraví.

Ve strukturované tabulkové podobě představují vrstvy 4D až 8D další atributy ve formě sloupců s vyplněnými hodnotami (obrázek 5.3-5), které jsou přidány k již vyplněným atributům 3D -modelu, jako je název, kategorie, typ a objemové charakteristiky. Hodnoty ve vrstvách atributů 6D, 7D a 8D obsahují další textové a číselné údaje, jako je procento recyklace, uhlíková stopa, záruční doba, cyklus výměny, datum instalace, bezpečnostní protokoly atd.

ID	Type Name	Width	Length	Recyclability	Carbon Footprint	Warranty Period	Replacement Cycle	Maintenance Schedule	Installation Date	Wellbeing Factors	Safety Protocols
W-NEW	Window	120 cm	-	90%	1622 kgCO ₂ e	8 years	20 years	Annual	-mon	XYZ Windows	ISO 45001
W-OLD1	Window	100 cm	140 cm	90%	1522 kgCO ₂ e	8 years	15 years	Biannual	08/22/2024	XYZ Windows	OSHA Standard
W-OLD2	Window	110 cm	160 cm	90%	1522 kgCO ₂ e	-	15 years	Biannual	08/24/2024	????	OSHA Standard
D-122	Door	90 cm	210 cm	100%	1322 kgCO ₂ e	15 years	25 years	Biennial	08/25/2024	Doors Ltd.	OSHA Standard

Obrázek 5.3-5 6D -8D přidávají atributové vrstvy do datového informačního modelu, který již obsahuje geometrické a objemové atributy z 3D -modelu.

Pro naše nové okno (obrázek 4.4-1) může mít prvek s identifikátorem W-NEW (obrázek 5.3-5) následující atributy 3D -8D:

3D -atributy - geometrické informace získané ze systémů CAD:

- "Type name" - element "Window"
- "Šířka" - 120 cm
- Kromě toho můžete přidat body "Bounding Box" prvku nebo jeho "geometrie BREP / MESH" jako samostatný atribut.

Atributy 6D - environmentální udržitelnost:

- Míra recyklace 90 procent
- "Uhlíková stopa" - 1 622 kg CO₂

Atributy 7D - údaje pro správu objektů:

- "Záruční doba" - 8 let
- "Náhradní cyklus" je 20 let starý
- "Údržba" - vyžaduje se každoročně

Atributy 8D - zajištění bezpečného užívání a provozu budov:

- Okno "Installed" - společnost "XYZ Windows"
- "Bezpečnostní norma" - v souladu s ISO 45001

Všechny parametry zaznamenané v databázi nebo souboru dat (obr. 5.3-5) potřebují odborníci v různých odděleních pro seskupování, vyhledávání nebo výpočty. Takový vícerozměrný atributový popis objektů projektu poskytuje úplný obraz o jejich životním cyklu, provozních požadavcích a mnoha dalších aspektech nezbytných pro návrh, výstavbu a provoz projektu.

Odhad emisí CO₂ a výpočet emisí oxidu uhličitého ze stavebních projektů

Vedle tématu udržitelnosti stavebních projektů ve fázi 6D (obrázek 5.3-5) se moderní stavebnictví zaměřuje na environmentální udržitelnost projektů, kde se jedním z klíčových aspektů stává hodnocení a minimalizace emisí oxidu uhličitého CO₂, které vznikají během fází životního cyklu projektu (např. výroba a montáž).

Odhad a výpočet emisí uhlíku ze stavebních materiálů je proces, při kterém se celkové emise uhlíku stanoví vynásobením objemových charakteristik prvku nebo skupiny prvků použitych v projektu vhodným faktorem emisí uhlíku pro danou kategorii.

Zohlednění emisí uhlíku při posuzování stavebních projektů jako součásti širších kritérií ESG (environmentální, sociální a správní) přidává do analýzy novou úroveň složitosti. To je důležité zejména pro klienta-investora při získávání příslušných certifikací, jako je LEED® (Leadership in Energy and Environmental Design), BREEAM® (Building Research Establishment Environmental Assessment Method) nebo DGNB® (Deutsche Gesellschaft für Nachhaltiges Bauen). Získání některé z těchto certifikací může výrazně zvýšit prodejnost nemovitosti, zjednodušit uvedení do provozu a zajistit splnění požadavků nájemců orientovaných na udržitelnost (ESG). V závislosti na požadavcích projektu lze využít také certifikáty HQE (Haute Qualité Environnementale, francouzský standard pro ekologické budovy), WELL (WELL Building Standard, zaměřený na zdraví a komfort uživatelů) a GRESB (Global Real Estate Sustainability Benchmark)

Environmentální, sociální a správní řízení **ESG** (environmental, social and governance) je široký soubor zásad, které lze použít k internímu i externímu hodnocení správy a řízení podniku a jeho sociálního a environmentálního dopadu.

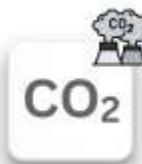
ESG, původně vyvinutý na počátku roku 2000 finančními fondy, aby investorům poskytl informace o širokých environmentálních, sociálních a správních kritériích, se vyvinul v klíčový ukazatel pro hodnocení společnosti i projektů, včetně stavebních. Podle průzkumů významných poradenských firem se environmentální, sociální a správní aspekty (ESG) stávají nedílnou součástí stavebnictví.

Podle publikace EY (2023) "The Path to Carbon Neutrality" (Cesta k uhlíkové neutralitě) společnosti, které aktivně uplatňují principy ESG, nejenže snižují dlouhodobá rizika, ale také zvyšují efektivitu svých obchodních modelů, což je obzvláště důležité při globální transformaci trhu [103]. Zpráva společnosti PwC o povědomí o ESG uvádí, že povědomí společností o významu ESG -faktorů se pohybuje od 67% do 97%, přičemž většina organizací považuje tyto trendy za klíčové pro budoucí udržitelnost [104] a že podniky z velké části zaznamenávají značný tlak ze strany zainteresovaných stran na integraci principů ESG.

Integrace principů ESG -principů do stavebních projektů tak nejen přispívá k získání mezinárodních certifikátů udržitelnosti, jako jsou LEED, BREEAM, DGNB, ale také zajišťuje dlouhodobou udržitelnost a konkurenčeschopnost firem v oboru.

Jedním z nejvýznamnějších faktorů, které ovlivňují celkovou uhlíkovou stopu stavebního projektu, jsou fáze výroby a logistiky stavebních materiálů a komponentů. Materiály používané na stavbě mají často rozhodující vliv na celkové emise CO₂, zejména v počátečních fázích životního cyklu projektu - od těžby surovin až po jejich dodání na staveniště.

Výpočet emisí podle kategorie nebo typu stavebního prvku vyžaduje použití referenčních emisních faktorů uhlíku, které odrážejí množství CO₂ vyprodukovaného při výrobě různých materiálů. Mezi tyto materiály patří beton, cihly, recyklovaná ocel, hliník a další. Tyto hodnoty se obvykle získávají z renomovaných zdrojů a mezinárodních databází, jako je například UK ICE 2015 (Inventory of Carbon and Energy) a US EPA 2006 (U.S. Environmental Protection Agency) [105]. Následující tabulka (obrázek 5.3-6) shrnuje základní emisní faktory pro řadu běžných stavebních materiálů. U každého materiálu jsou uvedeny dva klíčové parametry: měrné emise CO₂ (v kilogramech na kilogram materiálu) a přepočítací koeficienty objemu na hmotnost (v kilogramech na metr krychlový), které jsou nezbytné pro začlenění výpočtů do návrhového modelu a propojení se seskupením údajů QTO.



Carbon Emitted in Production		UK ICE Database (2015) USEPA (2006)	UK ICE Database (2015) USEPA (2006)	Coefficient m³ to kg
Material	Abbreviated	Process Emissions (kg CO2e/ kg of product) (K1)	Process Emissions (kg CO2e/ kg of product) (K2)	
Concrete	Concrete	0.12	0.12	2400
Concrete block	Concrete_block	0.13**	0.14	2000
Brick	Brick	0.24	0.32	2000
Medium density fiberboard (MDF)	MDF	0.39*	0.32	700
Recycled steel (avg recy content)	Recycled_steel	0.47	0.81	7850
Glass (not including primary mfg.)	Glass	0.59	0.6	2500
Cement (Portland, masonry)	Cement	0.95	0.97	1440
Aluminum (virgin)	Aluminum	12,79	16,6	2700

Obrázek 5.3-6 Množství uhlíku emitovaného při výrobě různých stavebních materiálů podle databáze britské ICE a americké EPA.

Pro výpočet celkových emisí CO₂ pro projekt je stejně jako u výpočtů 4D a 5D nutné určit objem atributů každé skupiny zařízení. To lze provést pomocí nástrojů kvantitativní analýzy (QTO) získáním atributových objemů v metrech krychlových, jak je podrobně popsáno v části o odběru množství. Tyto objemy se pak vynásobí příslušnými koeficienty pro atribut "Emise CO₂ z procesů" každé skupiny materiálů.

- ❷ Automaticky vytáhneme tabulku objemů podle typu prvku z projektu CAD (BIM) seskupením všech dat projektu, jak již bylo provedeno v předchozích kapitolách. K provedení tohoto úkolu využijme program LLM.

Prosím, seskupte tabulku DataFrame z projektu CAD (BIM) podle parametru sloupce "Název objektu" (nebo "Typ") a zobrazte počet položek v každé skupině a shrňte parametr "Objem" pro všechny položky daného typu. ↴

■ Odpověď společnosti LLM :

The screenshot shows a Jupyter Notebook environment. At the top, there are icons for 'Create in LLM Chat' and several AI models (Hugging Face, Qwen, StarCoder, and GPT-4). Below the header, there are three boxes: 'Data' (containing a CSV file icon), 'f(x)' (containing a function icon), and 'Output' (containing a table icon). The main area contains a code cell with the following Python code:

```

1 import pandas as pd
2
3 # Import Projekt as CSV tain DataFrame
4 df = pd.read_csv('your_dataframe.csv')
5
6 # Group, count, and sum 'Volume' in one line
7 result = df.groupby('ObjectType')['Volume'].agg(Count='size', Sum_of_Volumes='sum')

```

Below the code cell is a table titled 'Output' showing the results of the grouped data:

ObjectType	Amount	Sum of Volumes, m ³
IFC_Vebo_staalatei_L150B_90_x_150	31	412
IFC_Vebo_staalatei_L200B_90_x_200	4	80
IFC_betonlatei_200x250_200_x_250	4	68
IFC_betonopstort_150x315_150_x_315	12	152
M_Curtain_Wall_Dbl_Glass	6	0.42
M_Curtain_Wall_Sgl_Glass	12	0.33

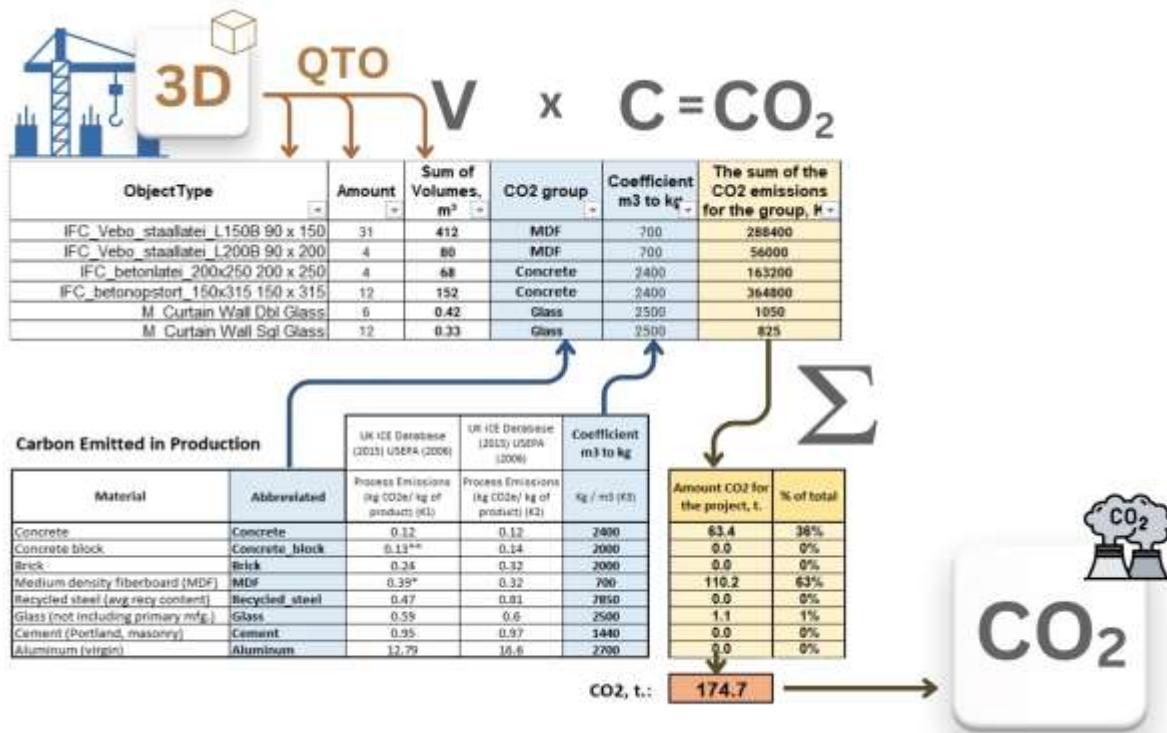
At the bottom left of the interface, there are buttons for 'Run in IDE' and other notebook controls.

Obrázek 5.3-7 Vygenerovaný kód v LLM nám seskupil entity projektu podle typu (ObjectType) se souhrnným atributem "Volume".

Chcete-li automatizovat výpočet celkových emisí CO₂ pro celý projekt, stačí nastavit automatické mapování dat v tabulce nebo ručně přiřadit typy prvků (obrázek 5.3-7) k odpovídajícím typům materiálů (obrázek 5.3-6) z tabulky emisních faktorů. Hotovou tabulku s emisními faktory a vzorcí, stejně jako kód pro načtení objemů z formátů CAD (BIM) a automatické stanovení CO₂ najdete na GitHubu po vyhledání "CO₂_calculating-the-embodyied-carbon. DataDrivenConstruction." [106].

Integrace dat po seskupení prvků QTO z databáze CAD tak umožňuje automatický výpočet emisí oxidu uhličitého (obr. 5.3-8) pro různé varianty návrhu. To umožňuje analyzovat vliv různých materiálů v různých variantách a vybrat pouze ta řešení, která splňují požadavky zákazníka na emise CO₂ pro získání konkrétního certifikátu při uvedení budovy do provozu

Odhad emisí CO₂ vynásobením faktorů objemem seskupených prvků projektu je typickým příkladem úkolu v procesu získávání hodnocení ESG (např. certifikace LEED) pro stavební podnik.

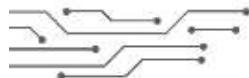


Obrázek 5.3-8 Integrace skupin QTO z databází CAD zajišťuje přesnost a automatizaci při odvozování odhadů konečných emisí CO₂.

Podobně můžeme definováním objemů skupin prvků provádět výpočty pro kontrolu a logistiku materiálu, sledování a řízení kvality, energetické modelování a analýzu a řadu dalších úloh, abychom získali nový stav atributů (parametr v tabulce) jak pro jednotlivé skupiny prvků, tak pro celý projekt.

Pokud se počet takových výpočtů ve firmě začne zvyšovat, vystane otázka, zda je třeba tyto výpočty automatizovat a implementovat výsledky výpočtů do firemních procesů a systémů správy dat.

Vzhledem ke složitosti kompletního řešení zadávají střední a velké společnosti ve stavebnictví tuto automatizaci společně s vyvíjejícími systémy ERP (nebo PMIS). Vývojové společnosti vytvářejí pro velké klienty jeden komplexní modulární systém pro správu mnoha různých informačních vrstev, včetně kalkulací materiálu a zdrojů.



KAPITOLA 5.4.

STAVEBNÍ ERP A PMIS SYSTÉMY

Stavební ERP -systémy na příkladu výpočtů a odhadů

Modulární systémy ERP integrují různé atributové (informační) vrstvy a datové toky do jediného komplexního systému, který umožňuje projektovým manažerům synchronizovaně řídit zdroje, finance, logistiku a další aspekty projektu v rámci jediné platformy. Stavební systém ERP funguje jako "mozek" stavebních projektů, zjednodušuje opakující se procesy prostřednictvím automatizace a zajišťuje transparentnost a kontrolu v celém procesu výstavby.

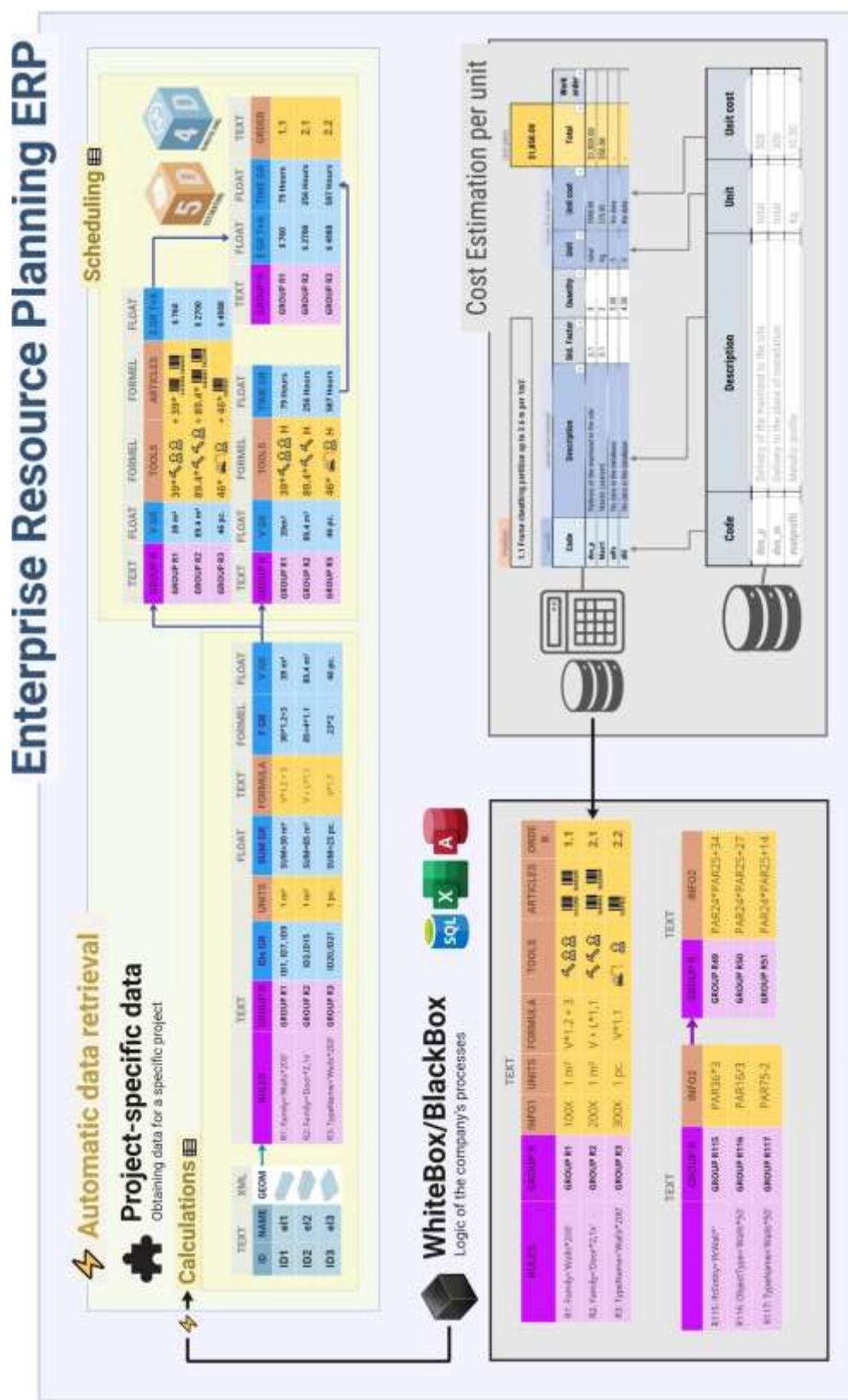
Stavební ERP -systémy (Enterprise Resource Planning) jsou komplexní softwarová řešení určená k řízení a optimalizaci různých aspektů stavebního procesu. Jádrem stavebních ERP systémů jsou moduly pro řízení kalkulace nákladů a plánování, což z nich činí důležitý nástroj pro efektivní plánování zdrojů.

Moduly ERP -systémy umožňují uživatelům strukturovaně zadávat, zpracovávat a analyzovat data pokrývající různé aspekty projektu, které mohou zahrnovat evidenci materiálových a mzdových nákladů, využití zařízení, řízení logistiky, lidské zdroje, kontakty a další stavební činnosti.

Jedním z funkčních bloků systému je modul automatizace obchodní logiky - BlackBox/WhiteBox, který plní roli řídicího centra procesu.

BlackBox /WhiteBox umožňuje specialistům používajícím ERP -systém flexibilně spravovat prostřednictvím přístupových práv různé aspekty podnikání, které již byly předem nakonfigurovány jinými uživateli nebo správci. V kontextu systémů ERP se termíny *BlackBox* a *WhiteBox* vztahují k úrovní transparentnosti a kontrolovatelnosti vnitřní logiky systému:

- **BlackBox** ("černá skříňka") - uživatel komunikuje se systémem prostřednictvím rozhraní bez přístupu k vnitřní logice provádění procesu. Systém provádí výpočty sám na základě předem definovaných pravidel, která jsou před koncovým uživatelem skryta. Ten zadá data a získá výsledek, aniž by věděl, jaké atributy nebo koeficienty byly uvnitř použity.
- **WhiteBox** ("bílý rámeček") - logika procesu je k dispozici pro prohlížení, přizpůsobení a úpravy. Pokročilí uživatelé, správci nebo integrátoři mohou ručně definovat algoritmy zpracování dat, pravidla výpočtu a scénáře interakce mezi subjekty projektu.



Obr. 5.4-1 Architektura stavebního ERP -systému, pro získání odhadů a harmonogramů prací ručním vyplněním objemových atributů.

Příkladem může být situace, kdy zkušený uživatel nebo správce nastaví pravidlo: které atributy v odhadu se mají mezi sebou násobit nebo seskupit podle určité vlastnosti a kam se má konečný výsledek zaznamenat. Později méně vyškolení odborníci, například odhadci, jednoduše načtou nová data do ERP prostřednictvím uživatelského rozhraní - a získají hotové odhady, harmonogramy nebo specifikace, aniž by museli psát kód nebo rozumět technickým detailům logiky.

V předchozích kapitolách byly v souvislosti s interakcí LLM diskutovány výpočetní a logické moduly. V prostředí ERP probíhají podobné výpočty a transformace v rámci modulů skrytých za rozhraním tlačítek a formulářů.

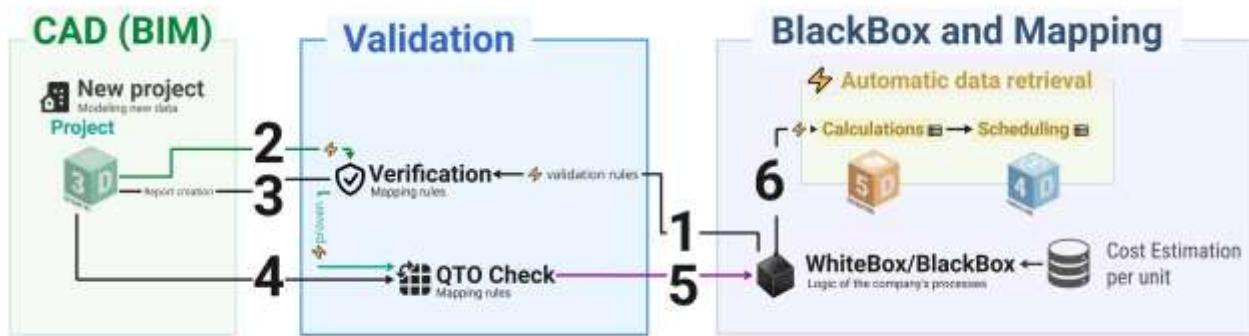
V následujícím příkladu (obr. 5.4-1) správce ERP -systému v modulu BlackBox /WhiteBox definoval pravidla pro párování atributů entit z odhadů s atributy pro seskupování QTO. Díky tomuto nastavení modulu BlackBox/WhiteBox (správcem nebo administrátorem) uživatel (odhadce nebo inženýr) ručním přidáním atributu množství nebo objemu prostřednictvím uživatelského rozhraní ERP automaticky obdrží hotové odhady a harmonogramy prací. Tímto způsobem se z procesů výpočtu a generování odhadů, o nichž se hovořilo v předchozích kapitolách, pomocí kódu uvnitř ERP stává poloautomatický dopravník.

Propojením tohoto poloautomatického procesu s objemovými atributy z modelů CAD (BIM) (obr. 4.1-13), například prostřednictvím načtení projektu CAD do modulu ERP, který je pro tento účel předem nakonfigurován, se tok dat změní v synchronizovaný mechanismus schopný samostatně a okamžitě aktualizovat hodnoty jednotlivých skupin prvků nebo celého projektu v reakci na jakékoli jeho změny ve fázi návrhu při načítání modelu CAD do ERP.

Aby bylo možné vytvořit automatizovaný tok dat (obr. 5.4-2) mezi systémy CAD (BIM) a ERP, je třeba strukturovaně definovat základní procesy a požadavky na data z databází modelů CAD (BIM), jak jsme již uvedli ve výše uvedené kapitole "Požadavky a zajištění kvality dat". Tento proces je v ERP rozdělen do podobných kroků:

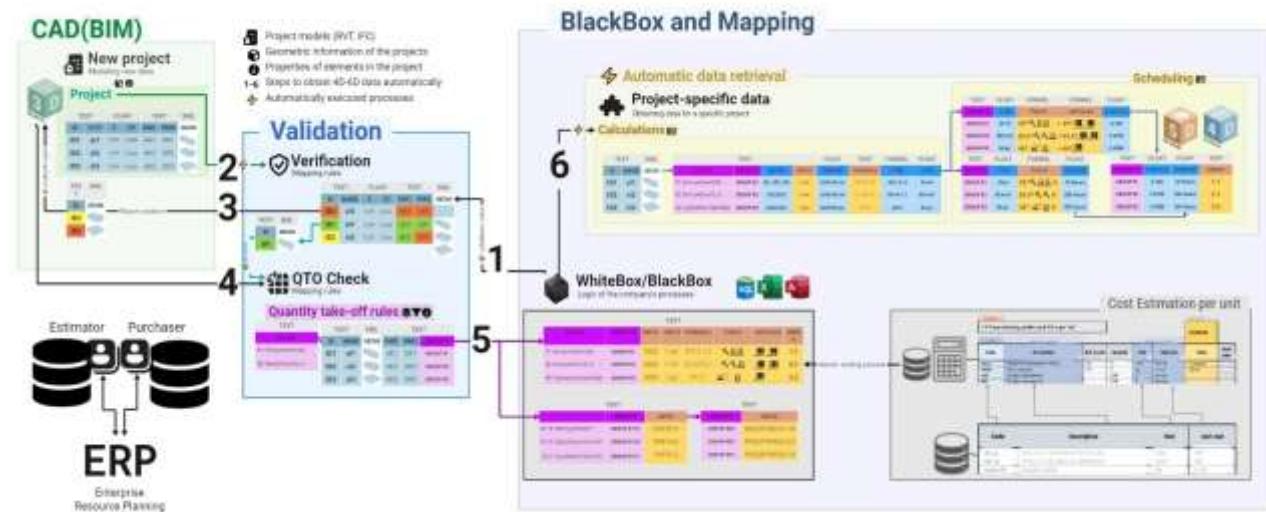
- **Vytvoření validačních pravidel (1)**, která hrají důležitou roli při zajišťování správnosti údajů vstupujících do systému ERP. Validační pravidla slouží jako filtry, které ověřují entity a jejich atributy a umožňují vstup do systému pouze těm položkám, které splňují požadavky. Více informací o ověřování a validaci se dozvíte v kapitole "Vytváření požadavků a ověřování kvality dat".
- Poté proběhne v ERP **ověřovací proces (2)**, který potvrdí, že všechny prvky projektových entit s jejich atributy a hodnotami byly vytvořeny správně a jsou připraveny pro další kroky zpracování.
- Pokud se vyskytnou problémy s neúplnými atributovými údaji, **je vygenerována zpráva (3)** a projekt je spolu s pokyny k opravě odeslán k revizi, dokud není připraven k další iteraci.
- Jakmile jsou údaje o projektu validovány a ověřeny, jsou použity v dalším modulu ERP **(4) k vytvoření tabulek množství (QTO)**, které vytvářejí atributy množství pro skupiny entit, materiály a zdroje podle předem vytvořených pravidel (WhiteBox/BlackBox).
- Seskupená data podle pravidel přiřazování nebo QTO jsou automaticky **integrována s výpočty (např. náklady a čas) (5)**.

- V posledním kroku ERP -systému uživatel vynásobením atributů rozsahu z tabulky QTO s atributy tabulek procesů (např. odhadovaných položek) **automaticky vygeneruje výsledky výpočtu** (6) (např. odhady nákladů , harmonogramy prací nebo emise CO₂) pro každou skupinu subjektů a pro projekt jako celek.

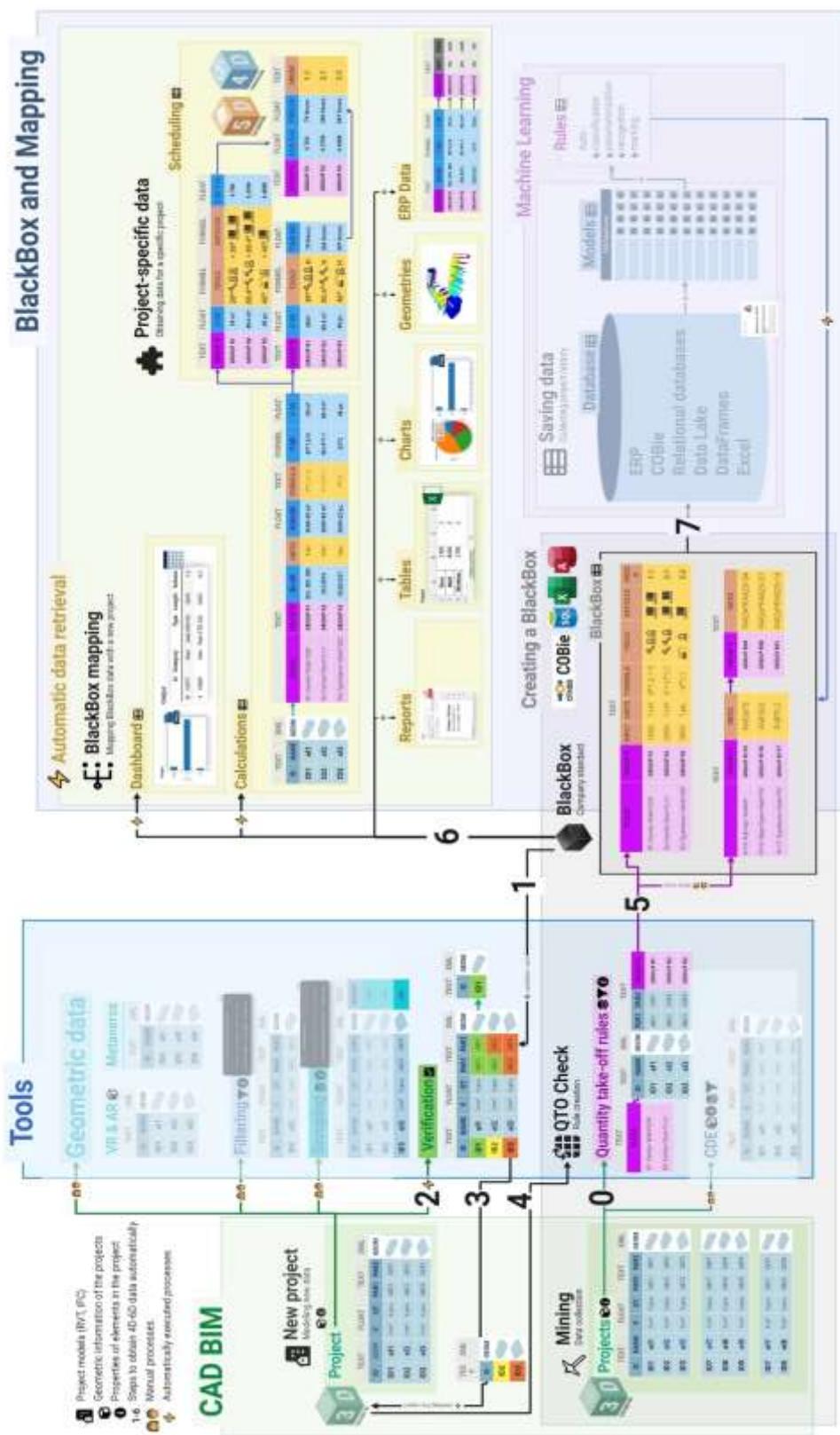


Obr. 5.4-2 Architektura stavebního ERP -systému s CAD (BIM), od tvorby validačních pravidel (1) až po automatický výpočet nákladů a harmonogramů prací (5-6).

V modulárním systému ERP jsou procesy integrovány pomocí softwaru, který obsahuje uživatelské rozhraní. Za rozhraním se nachází back-end, kde strukturované tabulky zpracovávají data prováděním různých operací, které byly předem nakonfigurovány manažerem nebo správcem. Výsledkem je, že uživatel díky předem definované a přizpůsobené automatizační logice (v modulech BlackBox /WhiteBox) dostává poloautomaticky připravené dokumenty, které splňují jeho úkoly



Obr. 5.4-3 ERP -systém pomáhá manažerům a uživatelům pohybovat se mezi odbornými tabulkami a vytváret nová data.



Obr. 5.4-4 ERP -systém je integrován s analytickými nástroji a automatizuje rozhodovací proces společnosti.

Podobně i procesy v ERP -systémech od počátku až po konečný výpočet (kroky 1-6 Obrázek 5.4-3) jsou řetězcem vzájemně propojených kroků, které v konečném důsledku zajišťují transparentnost, efektivitu a přesnost plánování.

Moderní stavební ERP -systémy obsahují nejen moduly pro výpočet nákladů a harmonogramu, ale také desítky dalších předkonfigurovaných modulů, které obvykle zahrnují správu dokumentů, sledování průběhu projektu, správu smluv, dodavatelský řetězec a logistiku, jakož i integraci s dalšími podnikovými systémy a platformami. Integrované analytické nástroje ERP umožňují uživatelům automatizovat tvorbu informačních panelů pro sledování klíčových ukazatelů výkonnosti projektu (KPI - key performance indicators). Tím je zajistěno centralizované a konzistentní řízení všech aspektů stavebního projektu se snahou spojit velké množství aplikací a systémů na jedné platformě.

V budoucnu se bude ERP -analytika používat v kombinaci se strojovým učením ke zvýšení přesnosti a optimalizaci procesu výpočtu budoucích atributů projektu. Data a atributy analyzované a shromážděné ze systémů ERP v rámci Big Data (obr. 5.4-4) budou v budoucnu tvořit základ pro tvorbu prediktivních modelů, které dokáží přesně předvídat potenciální zpoždění, rizika nebo například možné změny materiálových nákladů.

Jako alternativu k ERP, se ve stavebnictví často používá PMIS (Project Management Information System), systém řízení projektů určený pro detailní řízení úkolů na úrovni jednotlivých stavebních projektů.

PMIS: prostředník mezi ERP a stavbou.

Na rozdíl od ERP, který pokrývá celý řetězec podnikových procesů, se PMIS zaměřuje na řízení konkrétního projektu, sledování časového harmonogramu, rozpočtu, zdrojů a dokumentace.

PMIS (Project Management Information System) je software pro řízení stavebních projektů určený k plánování, sledování, analýze a vykazování všech aspektů projektu.

PMIS umožňuje spravovat dokumenty, harmonogramy, rozpočty a na první pohled se může zdát, že PMIS je duplicitním řešením k ERP, ale zásadní rozdíl je v úrovni řízení:

- **ERP** je zaměřen na obchodní procesy společnosti jako celku: náklady, zakázky, nákupy, lidské zdroje a řízení zdrojů na podnikové úrovni.
- **PMIS** se zaměřuje na řízení jednotlivých projektů, zajišťuje podrobné plánování, řízení změn, podávání zpráv a koordinaci účastníků.

V mnoha případech se jedná o ERP -systémy, které již mají dostatečnou funkčnost, a implementace PMIS se stává spíše otázkou pohodlí a preferencí společnosti. Mnoho dodavatelů a zákazníků nepoužívá PMIS proto, že by to bylo nutné, ale proto, že je to vynuceno dodavatelem nebo velkým zákazníkem, který chce agregovat data na určité platformě.

Je třeba zmínit, že v mezinárodní terminologii pro řízení stavebních projektů existují další samostatné populární pojmy, jako je PLM (Product Lifecycle Management) a EPC a EPC-M (Engineering, Procurement and Construction Management) - kontraktační metody ve stavebnictví.

Pokud společnost již používá ERP s moduly pro řízení projektů, může být zavedení PMIS zbytečným propojením, které duplikuje funkce. Pokud však procesy nejsou automatizované a data jsou roztríštěná, může být PMIS vhodnějším a snadno udržovatelným nástrojem.

Spekulace, zisk, izolovanost a nedostatek v ERP a PMIS

Navzdory vnější jednoduchosti rozhraní a postupů jsou stavební ERP a PMIS -systémy ve většině případů uzavřená a nepružná řešení. Takové systémy jsou obvykle dodávány jako předkonfigurovaný softwarový balík od jednoho dodavatele s omezeným přístupem k interním databázím a logice procesů.

Vývoj a řízení těchto systémů stále více přebírají dodavatelé CAD-(BIM-), protože jejich databáze obsahují informace požadované systémy ERP: kvantitativní a objemové atributy prvků projektu. Avšak místo toho, aby dodavatelé poskytovali přístup k těmto datům v otevřeném nebo strojově čitelném formátu, nabízejí pouze omezené uživatelské scénáře a uzavřenou logiku zpracování - předdefinovanou v rámci modulů BlackBox. To snižuje flexibilitu systému a brání jeho přizpůsobení specifickým podmínkám projektu.

Jedním z klíčových problémů digitálních procesů ve stavebnictví zůstává omezená transparentnost dat. Uzavřená architektura databází, nedostatečný přístup k úplným souborům atributů stavebních prvků, zaměření na *moduly* automatizace *BlackBox* a nedostatek otevřených rozhraní výrazně zvyšují rizika byrokracie dokumentů. Tato omezení vytvářejí úzká místa v rozhodovacím procesu, ztěžují ověřování informací a otevírají prostor pro skrývání dat nebo spekulace v rámci systémů ERP/PMIS. Uživatelé obvykle získávají pouze omezený přístup - ať už se jedná o ořezané rozhraní nebo částečné API - bez možnosti přímé interakce s primárními zdroji dat. To je obzvláště důležité, pokud jde o parametry automaticky generované z projektů CAD, jako jsou objemy, plochy a množství používané pro výpočty QTO.

V důsledku toho se mnoho stavebních společností namísto snahy o efektivitu prostřednictvím automatizace procesů, otevřených dat, snižování transakčních nákladů a vytváření nových obchodních modelů zaměřuje na řízení externích parametrů - manipulaci s faktory, korekčními faktory a metodami výpočtu ovlivňujícími náklady projektu v uzavřených platformách ERP/PMIS. To vytváří prostor pro spekulace, zkresluje skutečné výrobní náklady a snižuje důvěru mezi všemi účastníky stavebního procesu.

Ve stavebnictví se zisk tvoří jako rozdíl mezi výnosy z dokončeného projektu a variabilními náklady, které zahrnují projekt, materiál, práci a další přímé náklady přímo související s realizací projektu. Klíčovým faktorem ovlivňujícím hodnotu těchto nákladů však není pouze technologie nebo logistika, ale také rychlosť a přesnost kalkulací a kvalita manažerských rozhodnutí ve firmě.

Problém zhoršuje skutečnost, že ve většině stavebních společností zůstávají procesy kalkulace nákladů netransparentní nejen pro zákazníky, ale i pro samotné zaměstnance, kteří nejsou součástí oddělení odhadu nebo finančního oddělení. Taková uzavřenost napomáhá tomu, aby se v podniku vytvořila privilegovaná skupina specialistů - nositelů "finanční odbornosti", kteří mají výhradní právo upravovat atributy a korekční faktory v systémech ERP/PMIS. Tito zaměstnanci mohou spolu s vedoucími podniků fakticky řídit finanční logiku projektu.

Odhadci se v takových podmínkách mění ve "finanční žongléry", kteří balancují mezi maximalizací zisku společnosti a potřebou udržet konkurenčeschopnou cenu pro zákazníka. Zároveň se musí vyvarovat zjevných a hrubých manipulací, aby nepoškodili pověst společnosti. Právě v této fázi se stanoví koeficienty, které mají zakrýt nadhodnocené objemy nebo náklady na materiál a práce.

Výsledkem je, že hlavním schématem pro zvýšení efektivity a ziskovosti firem působících ve stavebnictví není automatizace a zrychlení rozhodovacích procesů, ale spekulace s cenami materiálů a prací (obr. 5.4-5). Nadhodnocování nákladů na práce a materiály se provádí "šedým" účetnictvím v uzavřených ERP /PMIS - systémech nadhodnocováním procent nad průměrné tržní ceny materiálů nebo objemů prací pomocí koeficientů (obr. 5.1-6), o kterých bylo pojednáno v kapitole "Sestavení kalkulací a výpočet nákladů na práce na základě zdrojové základny".

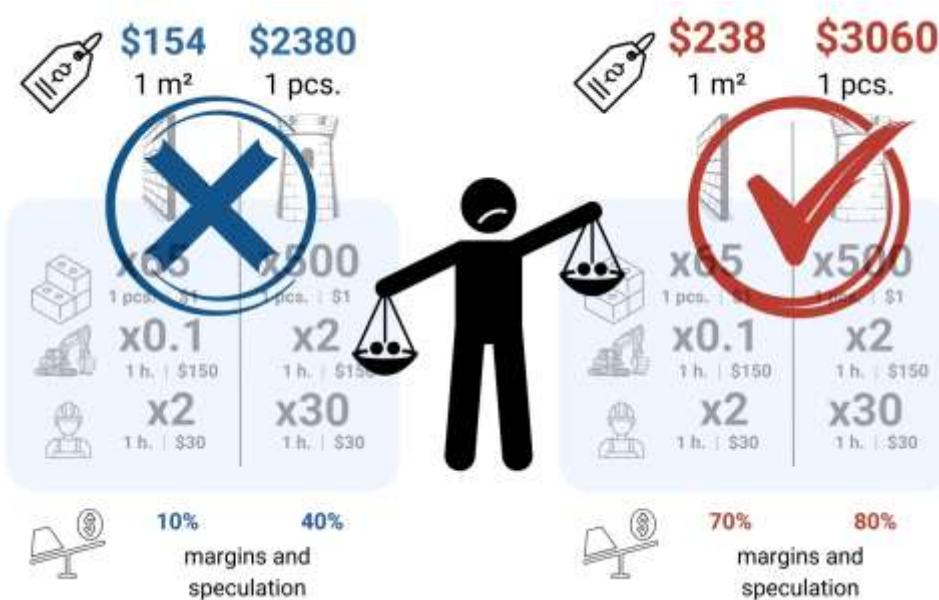
Výsledkem je, že zákazník obdrží kalkulaci, která neodráží skutečné náklady nebo rozsah prací, ale je derivátem mnoha skrytých interních koeficientů. Zároveň jsou subdodavatelé ve snaze vyhovět podhodnoceným sazbám stanoveným generálním dodavatelem často nuceni nakupovat levnější a nekvalitní materiály, což zhoršuje výslednou kvalitu stavby.

Spekulativní proces hledání zisků ze vzduchu nakonec poškozuje jak klienty, kteří dostávají nespolehlivé údaje, tak výkonné pracovníky, kteří jsou nuceni hledat další a další spekulativní modely. Čím větší je projekt, tím vyšší je míra byrokracie při řízení dat a procesů. Každý krok a každý modul často skrývá neprůhledné koeficienty a příplatky zakotvené ve výpočetních algoritmech a interních postupech. To nejen ztěžuje audit, ale také výrazně zkresluje finanční obraz projektu. U velkých stavebních projektů vedou tyto postupy často k několikanásobnému (někdy až desetinásobnému) navýšení konečných nákladů, zatímco skutečné objemy a náklady zůstávají mimo účinnou kontrolu zadavatele (obr. 2.1-3 Srovnání plánovaných a skutečných nákladů velkých infrastrukturních projektů v Německu).

Podle zprávy společnosti McKinsey & Company *Imagining the Digital Future of Construction* (2016) jsou velké stavební projekty dokončovány v průměru o 20% později, než bylo plánováno, a až o 80% nad rámec rozpočtu [107].

Oddělení odhadu a rozpočtování se stávají nejstřeženějším článkem společnosti. Přístup k nim je přísně omezen i pro interní specialisty a vzhledem k uzavřené logice a databázovým strukturám není možné objektivně a nezkresleně posoudit efektivitu projektových rozhodnutí. Netransparentnost vede k tomu, že firmy jsou nuceny nikoli optimalizovat procesy, ale bojovat o přežití "kreativním" řízením

čísel a faktorů (obr. 5.3-1, obr. 5.1-6 - např. parametr "Bid. faktor").



Obr. 5.4-5 Spekulativní poměry na úrovni vypořádání jsou hlavním ziskem firem a uměním žonglovat mezi kvalitou práce a dobrou pověstí.

To vše zpochybňuje další používání uzavřených systémů ERP/PMIS ve stavebnictví. V souvislosti s digitální transformací a rostoucími požadavky na transparentnost ze strany zákazníků (obr. 10.2-3) je nepravděpodobné, že by realizace projektů v dlouhodobém horizontu zůstala závislá na proprietárních řešeních, která omezují flexibilitu, brání integraci a brzdí rozvoj podnikání.

A bez ohledu na to, jak výhodné může být pro stavební firmy pracovat s datovými sily a neprůhlednými daty v uzavřených databázích - budoucnost stavebnictví bude nevyhnutelně zahrnovat přechod na otevřené platformy, strojově čitelné a transparentní datové struktury a automatizaci založenou na důvěře. Tato transformace bude řízena shora - pod tlakem zákazníků, regulačních orgánů a společnosti, která stále více požaduje odpovědnost, udržitelnost, transparentnost a ekonomickou proveditelnost.

Konec éry uzavřených ERP /PMIS: stavebnictví potřebuje nové přístupy.

Používání objemných modulárních systémů ERP/PMIS -systémů, které se skládají z desítek milionů řádků kódu, nesmírně ztěžuje jakékoli změny v nich. V takovém případě se přechod na novou platformu za přítomnosti již předkonfigurovaných modulů pro daný podnik, desítek tisíc článků v databázích zdrojů (obr. 5.1-3) a tisíců hotových kalkulací (obr. 5.1-6) mění v nákladný a zdlouhavý proces. Čím více kódů a starší architektury - tím vyšší úroveň vnitřní neefektivity a každý nový projekt situaci jen zhorší. V mnoha firmách se migrace dat a integrace nových řešení stávají mnohaletými eposy doprovázenými neustálým přepracováváním a nekonečným hledáním kompromisů. Výsledkem je často návrat ke starým, známým platformám, a to i přes jejich omezení.

Jak zdůrazňuje německá zpráva Black Book [108] o systémových nedostatkích ve správě stavebních

dat, roztríštěnost informací a absence centralizovaného přístupu k jejich správě je hlavní příčinou neefektivity. Bez standardizace a integrace ztrácejí data svou hodnotu a stávají se spíše archivem než nástrojem řízení.

Hlavní příčinou ztráty kvality dat je nedostatečné plánování a kontrola stavebních projektů, což často vede k výraznému zvýšení nákladů. Kapitola "Zaostřeno: Explosy nákladů" v Černé knize analyzuje klíčové faktory, které k této nežádoucímu důsledkům přispívají. Patří mezi ně nedostatečné analýzy potřeb, chybějící studie proveditelnosti a nekoordinované plánování vedoucí k dodatečným nákladům, kterým bylo možné předejít.

Ve vyspělém ekosystému IT společnosti je výměna zastaralého systému srovnatelná s výměnou nosného sloupu u již postavené budovy. Nestačí jen odstranit starý a nainstalovat nový - je důležité to udělat tak, aby budova zůstala stabilní, stropy se nezřítily a všechny komunikace fungovaly i nadále. Právě v tom spočívá potíž: jakákoli chyba může mít vážné důsledky pro celý systém společnosti.

Přesto vývojáři velkých ERP produktů pro stavebnictví stále používají množství napsaného kódu jako argument ve prospěch své platformy. Na specializovaných konferencích lze stále slyšet věty typu: "Znovuvytvoření takového systému by trvalo 150 člověkoroků", přestože většina funkcí takových systémů skrývá databáze a poměrně jednoduché funkce pro práci s tabulkami, zabalené do speciálního pevného, uživatelského rozhraní. V praxi se objem kódu "150 člověkoroků" mění spíše v zátež než v konkurenční výhodu. Čím více kódu - tím vyšší náklady na podporu, tím obtížnější přizpůsobení novým podmínkám a tím vyšší vstupní práh pro nové vývojáře a klienty.

Mnoho modulárních stavebních systémů dnes připomíná těžkopádné a zastaralé "frankensteinovské konstrukce", u nichž může každá neopatrná změna vést k poruše. Každý nový modul zvyšuje složitost již tak přetíženého systému a mění jej v labyrinth, kterému rozumí jen několik specialistů, což ještě více ztěžuje jeho údržbu a modernizaci.

Složitost si uvědomují i samotní vývojáři, kteří se pravidelně pozastavují nad refactoringem - revizí architektury s ohledem na vznik nových technologií. Nicméně i když se refactoring provádí pravidelně, složitost nevyhnutelně roste. Architekti takových systémů si na rostoucí složitost zvyknou, ale pro nové uživatele a specialisty se stává nepřekonatelnou překážkou. V důsledku toho se veškerá odbornost soustředí v rukou několika vývojářů a systém přestává být škálovatelný. Z krátkodobého hlediska jsou takoví odborníci užiteční, ale z dlouhodobého hlediska se stávají součástí problému.

Organizace budou i nadále integrovat "malá" data s jejich velkými protějšky a je bláhové se domnívat, že jedna aplikace - jakkoli drahá nebo robustní - zvládne vše [109].

- Phil Simon, moderátor podcastu Conversations About Collaboration (Rozhovory o spolupráci).

Vyvstává oprávněná otázka: Opravdu potřebujeme tak těžkopádné a uzavřené systémy pro výpočet nákladů a časového harmonogramu prací ve formě tabulek, když v jiných odvětvích si s podobnými úkoly již dávno poradily analytické nástroje s otevřenými daty a transparentní logikou?

V současné době jsou ve stavebnictví stále žádané uzavřené modulární platformy, a to především kvůli specifikům účtování nákladů (obr. 5.1-7). Takové systémy se často používají k provozování šedých nebo neprůhledných schémat, což umožňuje skrýt skutečné náklady před zákazníkem. Nicméně s tím, jak odvětví digitálně dozrává, především zákazníci, a přechází do takzvané "uberizované éry", budou zprostředkovatelé, tedy stavební firmy se svými ERP, ztráct svůj význam v časových a nákladových kalkulacích. To navždy změní tvář stavebnictví. Více se dočtete v poslední části knihy a v kapitole "Stavebnictví 5.0: Jak vydělávat peníze, když už se nemůžete schovat".

Tisíce starších řešení nashromážděných za posledních 30 let, do jejichž vývoje byly investovány tisíce let, začnou rychle mizet. Přechod na otevřenou, transparentní a flexibilní správu dat je nevyhnutelný. Jedinou otázkou je, které společnosti se těmto změnám dokážou přizpůsobit a které zůstanou rukojmím starého modelu.

Podobná situace je i v oblasti nástrojů CAD (BIM -), jejichž data dnes vyplňují objemové parametry návrhových entit v systémech ERP/PMIS -. Původně byla myšlenka BIM (vyvinutá již v roce 2002 [110]) založena na konceptu jediné integrované databáze, ale v praxi dnes práce s BIM vyžaduje celou řadu specializovaného softwaru a formátů. To, co mělo zjednodušit projektování a řízení výstavby, se změnilo v další vrstvě proprietárních řešení, která komplikují integraci a snižují flexibilitu podnikání.

Další kroky: efektivní využívání údajů o projektech

V této části jsme si ukázali, jak se strukturovaná data stávají základem pro přesné výpočty nákladů a harmonogramu stavebních projektů. Automatizace procesů QTO, plánování a odhadu snižuje náklady na pracovní sílu a výrazně zvyšuje přesnost výsledků.

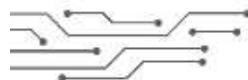
Pro shrnutí této části je vhodné zdůraznit hlavní praktické kroky, které vám pomohou aplikovat diskutované přístupy na vaše každodenní úkoly. Tyto přístupy jsou univerzální - jsou užitečné jak pro digitální transformaci podniku, tak pro každodenní práci odborníků zapojených do výpočtů:

- Automatizace rutinních výpočtů
 - Pokuste se najít standardní kalkulace nákladů na práci, na které se můžete při své práci odvolat.
 - Analyzujte, které metody se používají pro výpočet nákladů nebo kalkulaci prací nebo procesů na staveništi ve vaší zemi (obrázek 5.1-7).
 - Pokud pracujete se systémem CAD - seznamte se s funkcemi automatické extrakce specifikací a dat QTO ve vašem softwaru CAD (BIM-).
 - Použití nástroje LLM k psaní návrhu kódu pro automatizaci výpočtů.
- Vývoj vlastních nástrojů pro QTO

- Vytvoření skriptů nebo tabulek pro automatizaci počítání svazků
 - Standardizace kategorií a skupin prvků pro konzistentní přístup k hodnocení
 - zdokumentovat metodiku výpočtu, aby byla zajištěna reprodukovanost výsledků v nových projektech.
- Zapojte do své práce různé aspekty projektu.
- Pokud pracujete s modulárními systémy, snažte se své procesy vizualizovat nejen v podobě diagramů nebo grafů, ale také na úrovni dat - zejména v podobě tabulek.
 - Zvládnutí automatického slučování dat získaných z databází CAD s výpočty - pomocí kódu Pythonu s využitím seskupování, filtrování a agregace.
 - Vytvářet přehledné vizualizace skupin QTO pro prezentaci komplexních informací kolegům a zákazníkům.

Tyto kroky pomohou vytvořit udržitelný systém výpočtu založený na automatizaci a standardizaci údajů. Tento přístup zvýší přesnost a omezí rutinu každodenních výpočtů.

Následující kapitoly se zaměřují na technické aspekty produktů CAD - (BIM-) a na důvody, proč je stále obtížné integrovat databáze CAD do podnikových procesů. Pokud vás nyní nezajímá historie zavádění BIM ve stavebnictví, vývoj CAD nástrojů a technické aspekty práce s těmito technologiemi, můžete přejít rovnou k sedmé části knihy "Rozhodování založené na datech".



MAXIMÁLNÍ POHODLÍ S TIŠTĚNOU VERZÍ

Držíte v ruce bezplatnou digitální verzi knihy **Data-Driven Construction**. Pro pohodlnější práci a rychlý přístup k materiálům doporučujeme věnovat pozornost **tištěnému vydání**:



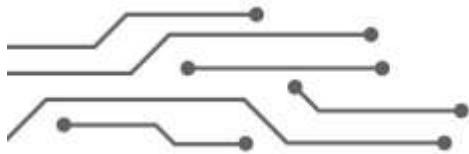
Vždy po ruce: kniha v tištěné podobě se stane spolehlivým pracovním nástrojem, který vám umožní rychle najít a použít správné vizualizace a diagramy v jakékoli pracovní situaci.

Vysoko kvalitní ilustrace: všechny obrázky a grafika v tištěném vydání jsou prezentovány v maximální kvalitě.

Rychlý přístup k informacím: snadná navigace, možnost dělat si poznámky, záložky a pracovat s knihou kdekoli.

Zakoupením plné tištěné verze knihy získáte praktický nástroj pro pohodlnou a efektivní práci s informacemi: možnost rychle využívat obrazové materiály při každodenních úkolech, rychle vyhledávat potřebná schémata a dělat si poznámky. Navíc svým nákupem podpoříte šíření otevřených znalostí.

Tištěnou verzi knihy si můžete objednat na adrese: datadrivenconstruction.io/books



VI ČÁST

CAD A BIM: MARKETING, REALITA A BUDOUCNOST PROJEKTOVÝCH DAT VE STAVEBNICTVÍ

Šestá část knihy představuje kritickou analýzu vývoje CAD a BIM -technologií a jejich dopadu na procesy správy dat ve stavebnictví. Sleduje historickou proměnu konceptu BIM od původní myšlenky integrované databáze až po současné marketingové konstrukce prosazované výrobci softwaru. Posuzuje se dopad proprietárních formátů a uzavřených systémů na efektivitu zpracování projektových dat a celkovou výkonnost stavebnictví. Podrobně jsou analyzovány problémy kompatibility různých CAD-systémů a obtíže jejich integrace s obchodními procesy stavebních společností. Jsou diskutovány současné trendy směrem ke zjednodušeným otevřeným datovým formátům, jako je USD, a jejich potenciální dopad na odvětví. Jsou představeny alternativní přístupy k získávání informací z uzavřených systémů, včetně technik reverzního inženýrství. Jsou analyzovány vyhlídky na využití umělé inteligence a strojového učení k automatizaci procesů navrhování a analýzy dat ve stavebnictví. Jsou formulovány prognózy vývoje projekčních technologií orientovaných na skutečné potřeby uživatelů, nikoliv na zájmy dodavatelů softwaru.

KAPITOLA 6.1.

VZNIK KONCEPCÍ BIM- VE STAVEBNICTVÍ

Původně tato šestá část, věnovaná CAD (BIM), nebyla do první verze knihy zařazena. Témata proprietárních formátů, geometrických jader a uzavřených systémů jsou příliš technická, přetížená detailem a zdánlivě zbytečná pro ty, kteří chtějí jen pochopit, jak pracovat s daty. Zpětná vazba a žádosti o doplnění vysvětlení do první verze knihy však ukázaly, že bez pochopení složitosti vnitřního fungování CAD systémů, geometrických jader, různorodosti formátů a nekompatibilních schémat ukládání stejných dat není možné skutečně pochopit, proč koncepty prosazované dodavateli často ztěžují práci s informacemi a brání přechodu na otevřené parametrické navrhování. Proto tato část zaujala ve struktuře knihy samostatné místo. Pokud pro vás CAD (BIM) není prioritou, můžete přeskočit rovnou na další část - "ČÁST VII: Rozhodování založené na datech, analytika, automatizace a strojové učení".

Historie vzniku BIM a otevřeného BIM jako marketingových konceptů dodavatelů CAD-.

S nástupem digitálních dat v 90. letech 20. století se počítačové technologie začaly používat nejen v obchodních procesech, ale také v procesech projektování, což vedlo ke vzniku pojmu, jako je CAD (computer-aided design systems) a později BIM (building information modelling)

Stejně jako každá inovace však nejsou konečným bodem vývoje. Koncepty jako BIM se staly důležitým milníkem v historii stavebnictví, ale dříve či později mohou ustoupit lepším nástrojům a přístupům, které budou lépe čelit výzvám budoucnosti.

Koncept BIM, který se objevil v roce 2002, zahlcený vlivem dodavatelů CAD a zmatený složitostí vlastní implementace, se dost možná nedožije svého třicátého výročí, podobně jako rocková hvězda, která se jasně rozzářila, ale rychle pohasla. Důvod je prostý: požadavky datových vědců se mění rychleji, než se jim dodavatelé CAD dokáží přizpůsobit.

Dnešní odborníci ve stavebnictví, kteří se potýkají s nedostatkem kvalitních dat, požadují interoperabilitu napříč platformami a přístup k otevřeným datům z projektů CAD- pro zjednodušení jejich analýzy a zpracování. Složitost dat CAD a nepřehledné zpracování dat CAD má negativní dopad na všechny účastníky stavebního procesu: projektanty, projektové manažery, stavební dělníky na stavbě a nakonec i na zákazníka.

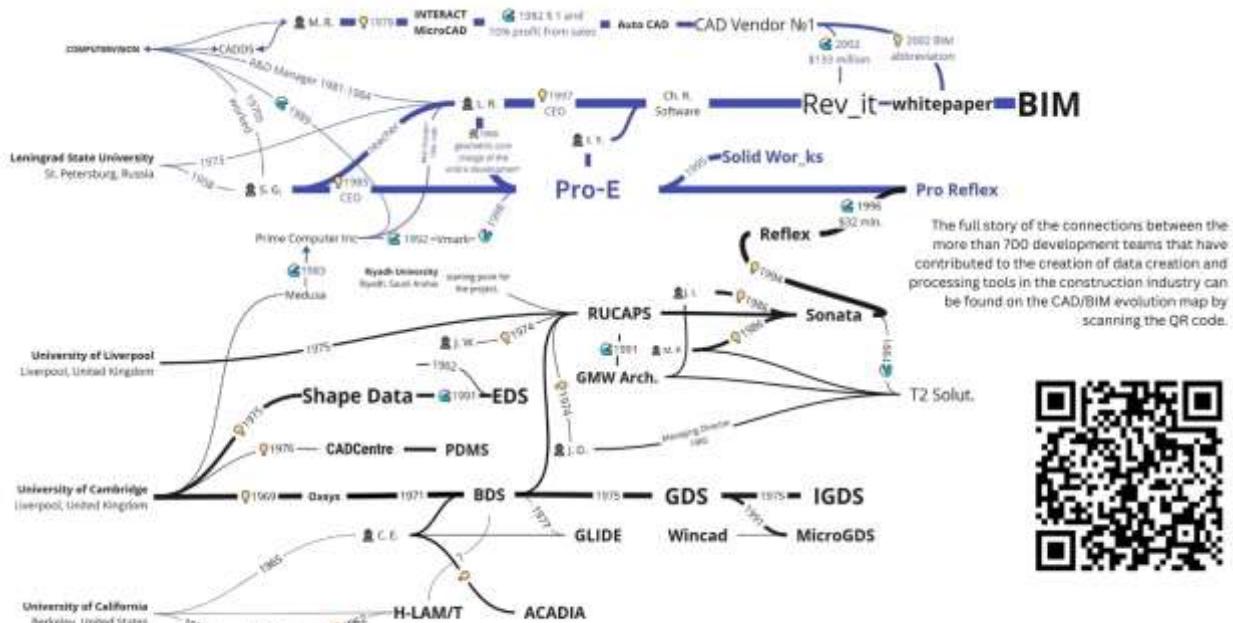
Místo plnohodnotné datové sady pro provoz dnes zákazník a investor dostávají kontejnery ve formátech CAD-, které vyžadují složitá geometrická jádra, porozumění datovým schématům, každoročně aktualizovanou dokumentaci API a specializovaný software CAD (BIM) pro práci s daty. Současně zůstává velká část návrhových dat nevyužita.

V dnešním světě projektování a výstavby vede složitost přístupu k datům CAD k nadměrnému

inženýrství při řízení projektů. Střední a velké společnosti pracující s daty CAD nebo vyvíjející BIM - řešení jsou buď nuteny udržovat úzké vztahy s dodavateli CAD řešení pro přístup k datům prostřednictvím API, nebo obcházet omezení dodavatelů CAD pomocí drahých SDK konvertorů k reverznímu inženýrství, aby získaly otevřená data [75].

Přístup založený na vlastních datech je zastaralý a již neodpovídá požadavkům dnešního digitálního prostředí. Budoucnost rozdělí společnosti na dva typy: ty, které budou efektivně využívat otevřená data, a ty, které trh opustí.

Koncept BIM (Building Information Modeling), se ve stavebnictví objevil s publikací jednoho z hlavních dodavatelů CAD - Whitepaper BIM [54] v roce 2002 a doplněný strojírenským konceptem BOM (Bills of Materials), vznikl z parametrického přístupu k tvorbě a zpracování projektových dat (obr. 6.1-1). Parametrický přístup k tvorbě a zpracování projektových dat byl jako jeden z prvních implementován v systému Pro-E pro strojírenské projektování (MCAD). Tento systém se stal prototypem [111mnoha moderních řešení CAD, včetně těch, která se dnes používají ve stavebnictví.



Obr. 6.1-1 Mapa historie konceptu BIM a podobných konceptů.

Novináři a konzultanti v oblasti AEC, kteří až do počátku roku 2000 propagovali nástroje CAD - prodejci, se od roku 2002 zaměřili na BIM Whitepaper. Právě Whitepaper BIM 2002-2004 a články publikované v letech 2002, 2003, 2005 a 2007 sehrály klíčovou roli při popularizaci konceptu BIM ve stavebnictví [112].

Informační modelování budov je strategie..... [název dodavatele CAD] pro aplikaci informačních technologií ve stavebnictví.

- BIM Whitepaper, 2002 [60].

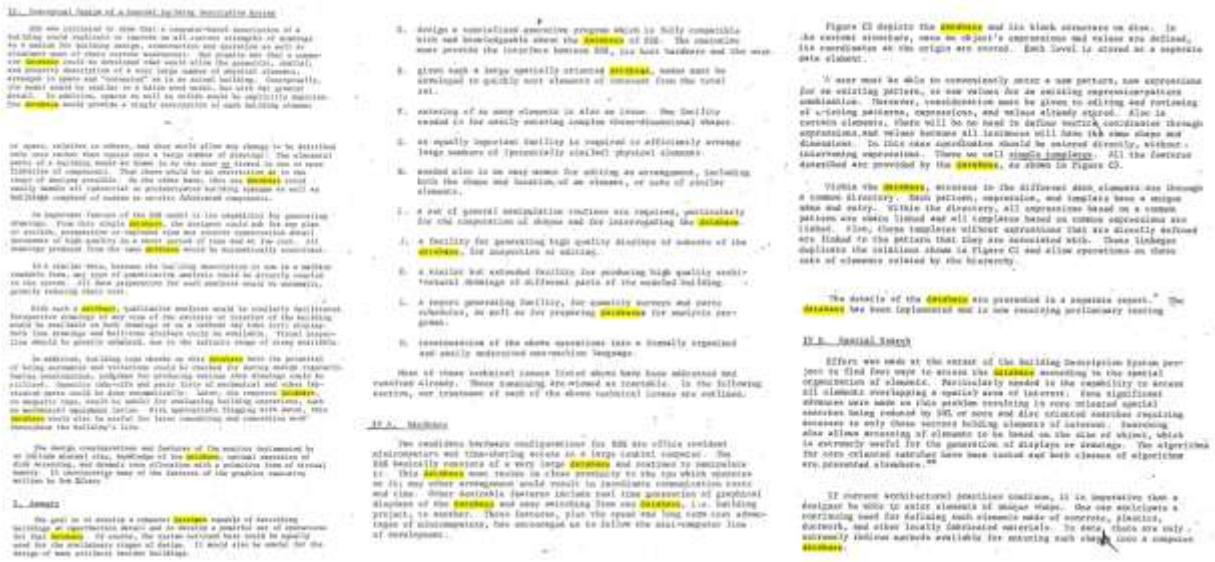
V polovině roku 2000 začali "výzkumníci" propojovat koncept BIM-, který v roce 2002 zveřejnil dodavatel CAD-, s dřívějšími vědeckými pracemi, jako je BDS Charlese Eastmana , které se staly základem pro systémy jako GLIDE, GBM, BPM, RUCAPS. Charles Eastman ve své průkopnické práci Building Description System (1974) položil teoretické základy moderního informačního modelování. Termín "databáze " se v jeho práci objevuje 43krát (obr. 6.1-2) - častěji než kterýkoli jiný, s výjimkou slova "budova".

Eastmanova klíčová myšlenka spočívala v tom, že všechny informace o budově - od geometrie až po vlastnosti prvků a jejich vzájemné vztahy - by měly být uloženy v jediné strukturované databázi. Z této databáze lze automaticky generovat a analyzovat výkresy, specifikace, výpočty a soulad s předpisy. Eastman výslově kritizoval výkresy jako zastaralý a nadbytečný způsob komunikace a poukazoval na duplicitu informací, problémy s aktualizací a nutnost ruční aktualizace při změnách. Místo toho navrhl jediný digitální model v databázi, kde se každá změna provede jednou a automaticky se promítnete do všech pohledů.

Je pozoruhodné, že Eastman ve své koncepci nestaví vizualizaci do centra pozornosti. V centru jeho systému byly informace: parametry, vztahy, atributy, možnosti analýzy a automatizace. Výkresy byly v jeho pojetí pouze jednou z forem zobrazení dat z databáze, nikoli primáním zdrojem informací o návrhu.

V první bílé knize o BIM od předního dodavatele CAD bylo slovní spojení "databáze " použito stejně často jako v BDS Charlese Eastmana - 23krát [60] sedmi stranách a bylo jedním z nejoblíbenějších slov v dokumentu po slovech "budova", "informace", "modelování" a "návrh". V roce 2003 se však termín "databáze" v podobných dokumentech objevuje pouze dvakrát [61] a koncem roku 2000 se téma databází z diskuse o datech pro navrhování prakticky vytratilo. Výsledkem bylo, že koncept "jednotné integrované databáze pro vizuální a kvantitativní analýzu" nebyl nikdy plně realizován.

Stavebnictví se tak dostalo od progresivního konceptu BDS Charlese Eastmana s důrazem na databáze a myšlenky Samuela Geisberga o automatické aktualizaci návrhových dat z databází ve strojírenském produktu Pro-E (předchůdce dnes ve stavebnictví používaných populárních CAD -řešení) až k současnemu marketingovému BIM, kde se o správě dat prostřednictvím databází téměř nemluví, přestože právě tato koncepce stála u zrodu původních teoretických



Obrázek 6.1-2 V konceptu BDS, který popsal Charles Eastman v roce 1974, bylo slovní spojení "Databáze" (zvýrazněno žlutě) použito 43krát.

BDS a podobné koncepty byly až do roku 2000 vyvíjeny spíše jako digitální databáze budov než jako vizualizační nástroj. BIM se v roce 2002 stal nástrojem pro navrhování, kde databáze ustoupila do pozadí. Co jsme ztratili při přechodu od BDS a podobných konceptů v 90. letech k BIM do poloviny roku 2010:

- Otevřené databáze: BDS a další podobné koncepty kladly důraz na analytiku, BIM na návrh.
- Flexibilita práce s daty: BDS kladl důraz na analýzu dat, BIM kladl důraz na procesy, které mají vycházet z nejasných dat.
- Transparentnost: BDS byl zamýšlen jako otevřená integrovaná databáze, zatímco dodavatelé CAD v BIM učinili své databáze zcela uzavřenými a již 20 let neúspěšně bojují proti nástrojům reverzního inženýrství, které otevírají proprietární formáty.

Za posledních 30 let neměli projektanti nikdy přístup k "integrované databázi" a po dvaceti letech marketingové euforie kolem nástrojů BIM si stavebnictví začíná uvědomovat důsledky této módní vlny.

Realita BIM: místo integrovaných databází - uzavřené modulární systémy.

Místo toho, aby se uživatelé systémů CAD - (BIM-) zaměřili na data, strukturovali je a integrovali do jednotných procesů, jsou nutni pracovat s roztríštěným souborem proprietárních řešení, z nichž každé si diktuje vlastní pravidla hry:

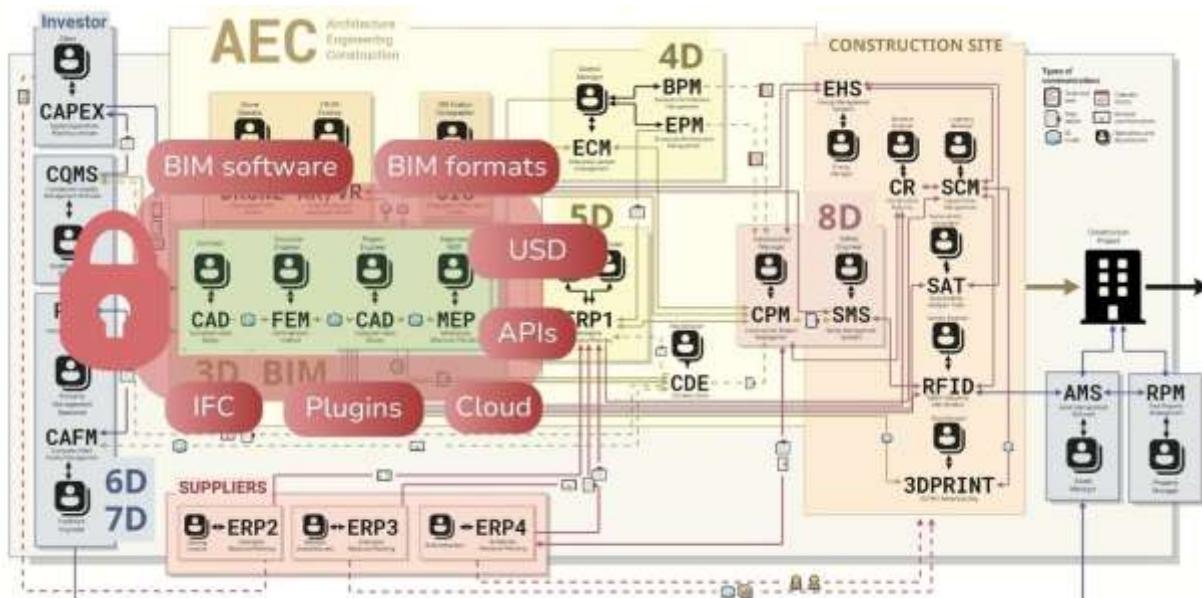
- **Jednotná databáze**, která byla předmětem první BIM Whitepaper, **zůstala mýtem**. Navzdory hlasitým prohlášením je přístup k datům stále omezený a rozdělený mezi uzavřené systémy.
- **BIM - modely se staly spíše uzavřeným ekosystémem** než nástrojem. Místo transparentní výměny informací jsou uživatelé nutni platit za předplatné a používat proprietární API.
- **Data patří projevcům, nikoli uživatelům**. Informace o projektech jsou uzamčeny v proprietárních formátech nebo cloudových službách, místo aby byly k dispozici v otevřených

a nezávislých formátech.

Konstruktéři a projektoví manažeři často nemají přístup k databázi CAD -systémů ani k formátu, ve kterém jsou uložena jejich vlastní projektová data. To znemožňuje rychlé ověření informací nebo formulaci požadavků na strukturu a kvalitu dat (obr. 6.1-3). Přístup k těmto datům vyžaduje celou řadu specializovaných programů propojených prostřednictvím API a plug-inů, což vede k nadměrné byrokratizaci procesů ve stavebnictví. Přitom tato data současně využívají desítky informačních systémů a stovky specialistů.

*Musíme být schopni všechna tato data [CAD (BIM)] digitálně spravovat, ukládat a prodávat software pro řízení životního cyklu a procesů, protože **na každého inženýra** [projektanta], který něco vytvoří [v softwaru CAD], **připadá deset lidí, kteří s těmito daty pracují**" [41].*

- Generální ředitel společnosti CAD - dodavatele, který vytvořil koncept BIM, 2005.



Obr. 6.1-3 CAD- (BIM-) databáze zůstávají jedním z posledních uzavřených systémů pro IT oddělení a správce dat v ekosystému stavebního podnikání.

Když je zřejmé, že BIM je spíše prostředkem komericializace databází než plnohodnotným nástrojem pro správu databází, vyvstává logická otázka: jak získat zpět kontrolu nad daty? Odpověď je použití otevřených datových struktur, kde se vlastníkem informací stává uživatel, nikoli dodavatel softwaru.

Uživatelé a vývojáři řešení ve stavebnictví, stejně jako jejich kolegové v jiných odvětvích, se nevyhnutelně odkloní od vágní terminologie dodavatelů softwaru, která převládala posledních 30 let, a

zaměří se na klíčové aspekty digitalizace - "data" a "procesy".

Ještě koncem 80. let byla klíčová oblast digitálního rozvoje ve stavebnictví prezentována jako otázka přístupu k datům a řízení projektových informací. Postupem času se však těžiště změnilo. Namísto rozvoje transparentních a přístupných přístupů k práci s daty se aktivně prosazoval formát IFC a otevřený koncept BIM jako pokusy odvést pozornost odborníků od témat správy projektových databází.

Vznik otevřeného formátu IFC ve stavebnictví

Takzvaný otevřený formát IFC (Industry Foundation Classes) je umístěn jako standard pro zajištění interoperability mezi různými systémy CAD (BIM -). Jeho vývoj probíhal v rámci organizací, které byly vytvořeny a řízeny hlavními dodavateli CAD. Na základě formátu IFC vyvinuly dvě společnosti CAD- v roce 2012 marketingový koncept OPEN BIM[63].

IFC (Industry Foundation Classes) je otevřený standard pro výměnu dat ve stavebnictví, který má zajistit interoperabilitu mezi různými systémy CAD - (BIM-).

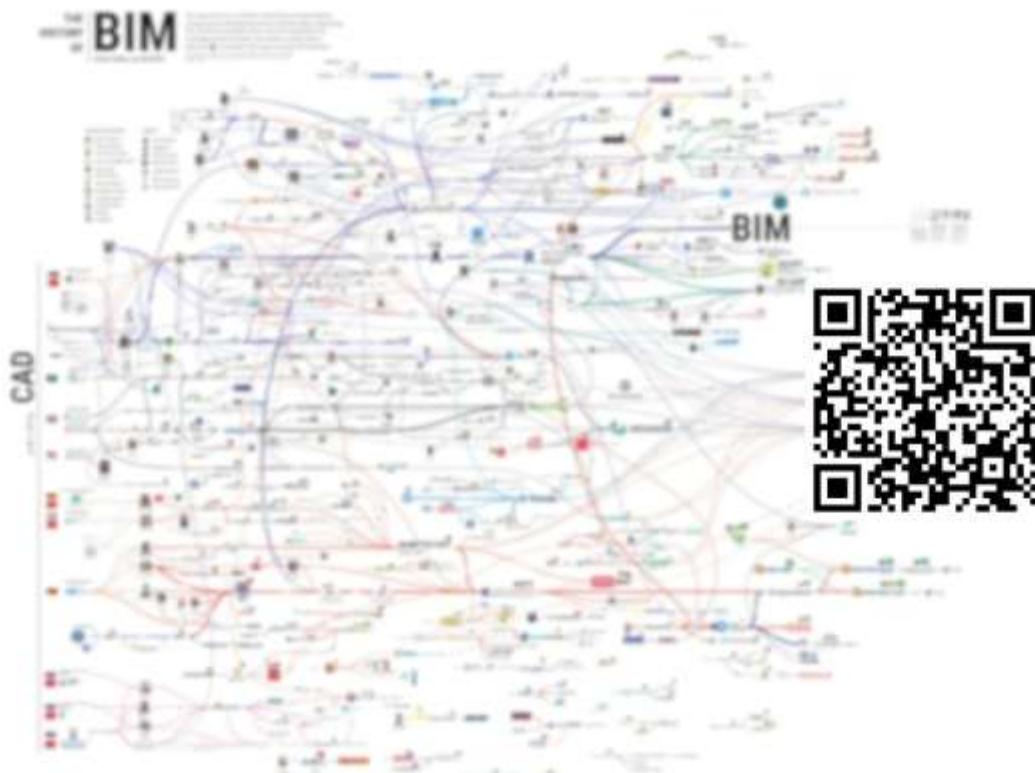
Koncept Open BIM - zahrnuje práci s informacemi z databází CAD a výměnu informací mezi systémy prostřednictvím otevřeného formátu pro výměnu dat CAD - IFC.

Program Open BIM je marketingová kampaň iniciovaná... [1 dodavatel CAD]... [2 dodavatel CAD] a dalšími společnostmi s cílem podpořit a usnadnit globální koordinovanou propagaci konceptu OPEN BIM v celém odvětví AEC, přičemž účastníci programu mají k dispozici jednotnou komunikaci a společnou značku.

- Z webových stránek dodavatele CAD, program OPEN BIM, 2012 [113].

Formát IFC byl koncem 80. let adaptován Technickou univerzitou v Mnichově ze strojírenského formátu STEP, později jej zaregistrovala významná konstrukční společnost a významný dodavatel CAD- a v roce 1994 vytvořila IAI (Industry Alliance for Interoperability) [114] (obr. 6.1-4). Formát IFC byl vyvinut k zajištění interoperability mezi různými systémy CAD a vycházel ze zásad stanovených ve strojírenském formátu STEP, který zase vznikl z formátu IGES vytvořeného v roce 1979 skupinou uživatelů a dodavatelů CAD s podporou NIST (The National Institute of Standards and Technology) a amerického ministerstva obrany [115].

Složitá struktura IFC, jeho úzká závislost na geometrickém jádru, stejně jako rozdíly v implementaci formátu různými softwarovými řešeními však vedly k mnoha problémům při jeho praktickém použití. S podobnými potížemi - ztrátou detailů, omezením přesnosti a nutností používat přechodné formáty - se dříve setkávali specialisté na strojírenství při práci s formáty IGES a STEP, z nichž IFC vznikl.



Obr. 6.1-4 Mapa vazeb mezi vývojovými týmy a produkty CAD (BIM) [116].

V roce 2000 vydává stejný dodavatel CAD, který zaregistroval formát IFC a vytvořil organizaci IAI (později bS), Whitepaper "Integrated Design and Manufacturing: Benefits and Rationale". [65]. Dokument zdůraznil důležitost zachování plné granularity dat při výměně mezi programy v rámci jednoho systému, aniž by se používaly neutrální formáty, jako je IGES, STEP [identický s IFC]. Místo toho bylo navrženo, aby aplikace měly přímý přístup k základní databázi CAD, aby se zabránilo ztrátě přesnosti informací.

V roce 2002 stejný dodavatel CAD kupuje parametrický produkt BOM (obr. 3.1-18, podrobněji na třetí části) a na jeho základě vytváří koncept BIM. V důsledku toho se při výměně dat stavebních projektů používají pouze uzavřené formáty CAD nebo formát IFC (STEP), o jehož omezeních psal v roce 2000 sám dodavatel CAD, který tento formát do stavebnictví přinesl.

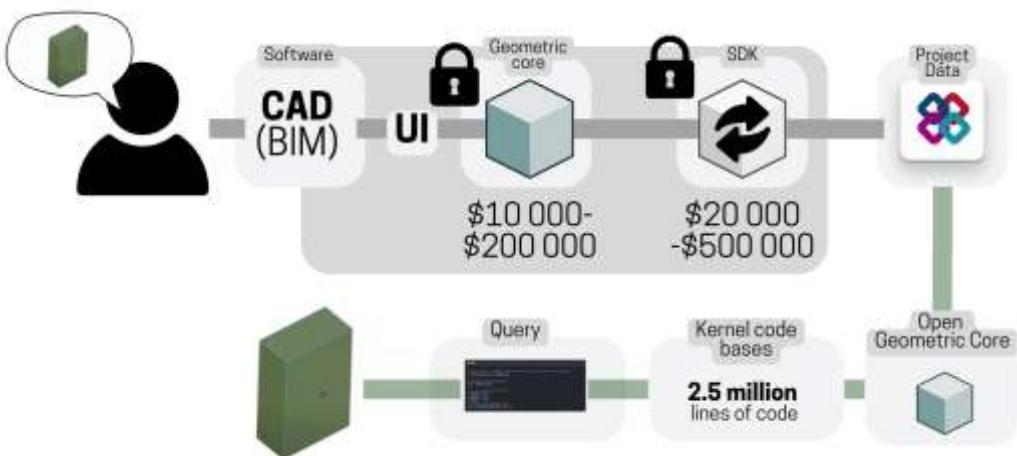
Podrobná historie interakce více než 700 vývojových týmů, které se podílely na tvorbě a zpracování stavebních dat, je uvedena v mapě "Vývoj CAD (BIM)" [116]. [116].

Otevřený formulář IFC se skládá z geometrického popisu konstrukčních prvků a popisu metainformací. Pro reprezentaci geometrie ve formátu IFC se používají různé metody, například CSG a Swept Solids: nicméně parametrická reprezentace BREP se stala hlavním standardem pro přenos geometrie prvků ve formátu IFC, protože tento formát je podporován při exportu z programů CAD- (BIM-) a umožňuje případné úpravy prvků při importu IFC zpět do programů CAD.

Problém s formátem IFC v závislosti na geometrickém jádru

Ve většině případů, kdy je geometrie v IFC definována parametricky (BREP), je nemožné vizualizovat nebo načítat geometrické vlastnosti, jako je objem nebo plocha entit projektu, pouze pomocí souboru IFC, protože pro práci s geometrií a její vizualizaci je v tomto případě nutné geometrické jádro (obrázek 6.1-5), které zpočátku chybí.

Geometrické jádro je softwarová komponenta, která poskytuje základní algoritmy pro vytváření, úpravy a analýzu geometrických objektů v CAD (CAD), BIM a dalších inženýrských aplikacích. Je zodpovědné za konstrukci 2D a 3D -geometrie a za operace s ní, jako jsou: booleovské operace, vyhlazování, průsečíky, transformace a vizualizace.

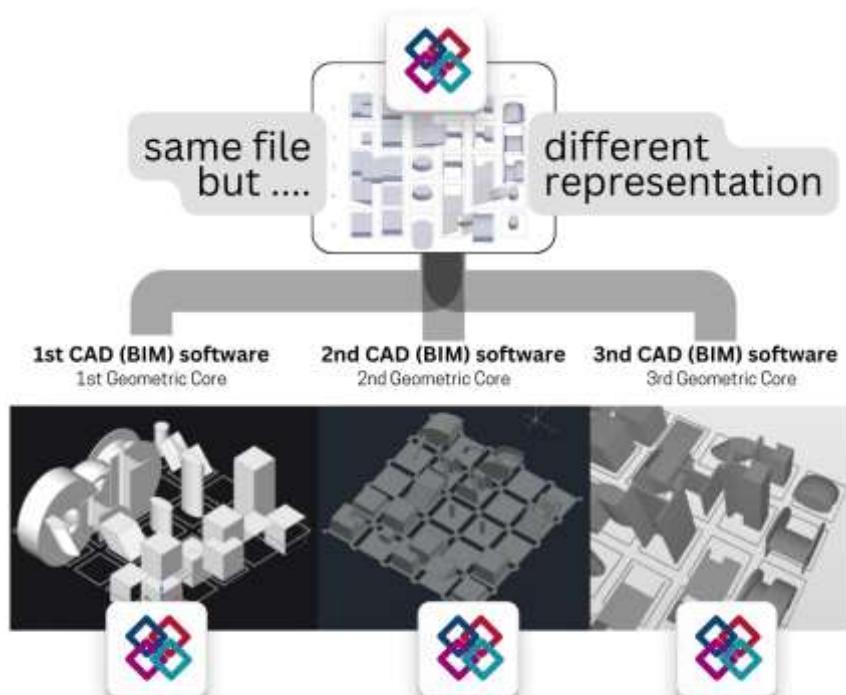


Obrázek 6.1-5 Vytváření geometrie prostřednictvím softwaru CAD- dnes probíhá prostřednictvím proprietárních geometrických jader a SDK, které často nejsou ve vlastnictví dodavatelů CAD.

Každý program CAD a všechny programy pracující s parametrickými formáty nebo formáty IFC mají vlastní nebo zakoupené geometrické jádro. A pokud s primitivními prvky ve formátu IFC -BREP nemohou být problémy a v programech s různými geometrickými jádry lze tyto prvky zobrazit podobně, ale kromě problémů s různými enginy geometrických jader, existuje dost prvků, které mají pro správné zobrazení svá specifika. Podrobně se touto problematikou zabývá mezinárodní studie "A reference study of IFC software support" publikovaná v roce 2019 [117].

Stejné standardizované datové sady přinášejí rozporuplné výsledky, bylo nalezeno jen málo společných vzorů a byly zjištěny vážné problémy s podporou standardu [IFC], pravděpodobně kvůli velmi vysoké složitosti standardního datového modelu. Částečně jsou zde na vině samotné standardy, které často ponechávají některé detaily nedefinované, s vysokým stupněm volnosti a různými možnými výklady. Umožňují vysokou složitost organizace a ukládání objektů, což neprospívá efektivnímu univerzálnímu porozumění, jedinečným implementacím a konzistentnímu modelování dat [117].

- Referenční studie softwarové podpory IFC, 2021



Source: Reference study of IFC software support: the GeoBIM benchmark 2019 — Part I <https://arxiv.org/pdf/2007.10951>

Obr. 6.1-6 Různá geometrická jádra poskytují různé reprezentace stejné geometrie popsané parametricky (podle [117]).

Správné pochopení "některých ustanovení" je k dispozici placeným členům speciálních organizací, které vyvíjejí IFC. V důsledku toho se ten, kdo chce získat přístup k důležitým poznatkům o určitých vlastnostech IFC, bude snažit spolupracovat s velkými dodavateli CAD-, nebo dosáhnout kvalitního posouzení vlastností vlastním výzkumem.

Narazíte na otázku týkající se importu a exportu dat prostřednictvím formátu IFC a zeptáte se kolegů prodejců: "Proč je tak v souboru IFC informace o parametrickém přenosu prostor? V otevřené specifikaci se o tom nic nepíše". Odpověď od "znejajších" evropských dodavatelů: "Ano, není to řečeno, ale je to povoleno".

- Z rozhovoru s vývojářem CAD 2021 [118].

IFC popisuje geometrii prostřednictvím parametrických primitiv, ale neobsahuje vložené jádro - jeho úlohu plní program CAD, který geometrii sestavuje prostřednictvím geometrického jádra. Jádro geometrie provádí matematické výpočty a definuje průsečíky, zatímco IFC pouze poskytuje data pro jejich interpretaci. Pokud IFC obsahuje nesprávné plochy, různé programy s různými geometrickými jádry je mohou v závislosti na jádru buď ignorovat, nebo vytvářet chyby.

V důsledku toho je pro práci s formátem IFC nutné zodpovědět hlavní otázku, na kterou je obtížné najít jednoznačnou odpověď - jaký nástroj, s jakým geometrickým jádrem použít, aby se dosáhlo kvality dat, kterou měl projekt původně v programu CAD, z něhož byl IFC získán?

Problémy s kvalitou dat a složitost formátu IFC neumožňují přímé využití návrhových dat pro automatizaci procesů, analýzu a zpracování dat, což často vede vývojáře k nevyhnutelné potřebě používat uzavřené CAD -řešení s "kvalitním" přístupem k datům[63], o čemž psal sám dodavatel, který IFC zaregistroval v roce 1994 [65].

Všechny zvláštnosti mapování a generování parametrů IFC v geometrickém jádře mohou realizovat pouze velké týmy vývojářů, kteří mají zkušenosti s prací s geometrickými jádry. Proto je současná praxe zvláštností a složitosti formátu IFC výhodná především pro dodavatele CAD- a má mnoho společného se strategií velkých dodavatelů softwaru "adoptovat, rozšířit, zničit", kdy rostoucí složitost standardu ve skutečnosti vytváří překážky pro malé hráče na trhu [94].

Strategií velkých dodavatelů v rámci takové strategie může být přizpůsobení otevřených standardů, přidání proprietárních rozšíření a funkcí, aby vytvořili závislost uživatelů na svých produktech a vytlačili tak konkurenci.

Formát IFC, který má být univerzálním mostem mezi různými systémy CAD- (BIM-), ve skutečnosti plní roli indikátoru problémů s kompatibilitou mezi geometrickými jádry různých platform CAD, podobně jako formát STEP, z něhož původně vznikl.

V důsledku toho je dnes úplná a kvalitní implementace ontologie IFC proveditelná pro velké dodavatele CAD, kteří mohou investovat značné prostředky do podpory všech entit a jejich mapování na vlastní interní geometrické jádro, což pro IFC jako standard neexistuje. Velcí dodavatelé mají také možnost mezi sebou koordinovat technické detaily prvků, které nemusí být dostupné ani nejaktivnějšímu účastníkovi organizací vyvíjejících formát IFC.

Pro malé nezávislé týmy a projekty s otevřeným zdrojovým kódem, které se snaží podporovat vývoj interoperabilních formátů, se absence vlastního geometrického jádra stává vážným problémem. Bez něj je prakticky nemožné zohlednit všechny různé jemnosti a nuance spojené s výměnou dat mezi platformami.

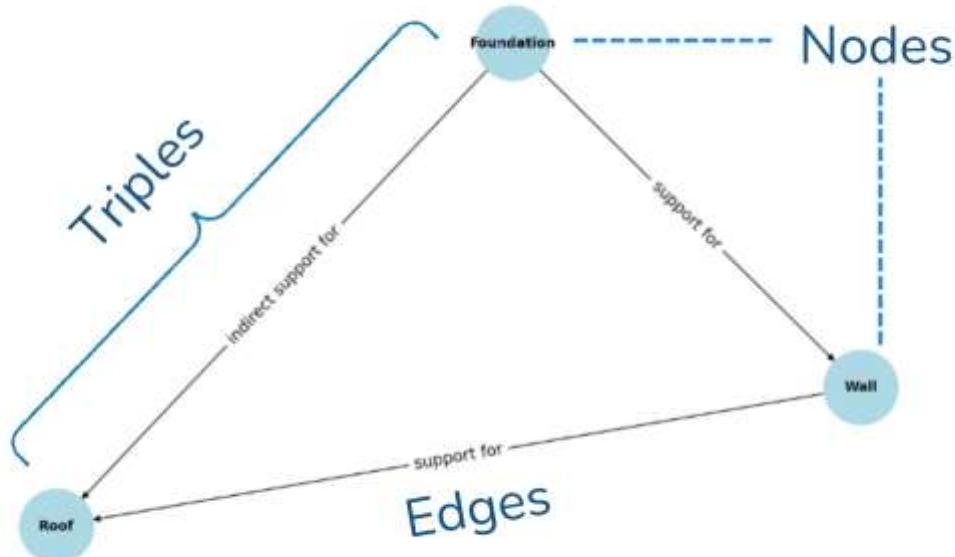
S rozvojem parametrického formátu IFC a otevřené koncepce BIM se ve stavebnictví zintenzivnily diskuse o úloze ontologie a sémantiky při správě dat a procesů.

Výskyt v konstrukci tématu sémantiky a ontologie

Díky myšlenkám sémantického internetu konce 90. let 20. století a úsilí organizací podílejících se na vývoji formátu IFC, se sémantika a ontologie staly jedním z klíčových prvků standardizace, o které se ve stavebnictví diskutuje do poloviny 20. století.

Sémantické technologie představují sjednocení, standardizaci a úpravu velkých polí heterogenních dat a realizaci komplexního vyhledávání.

Pro ukládání sémantických dat se používá jazyk OWL (Web Ontology Language), reprezentovaný jako grafy RDF -triplety (Resource Description Framework) (obr. 6.1-7). Jazyk OWL odkazuje na grafové datové modely, jejichž typy jsme podrobněji probrali v kapitole "Datové modely: datové vztahy a vztahy mezi prvky".



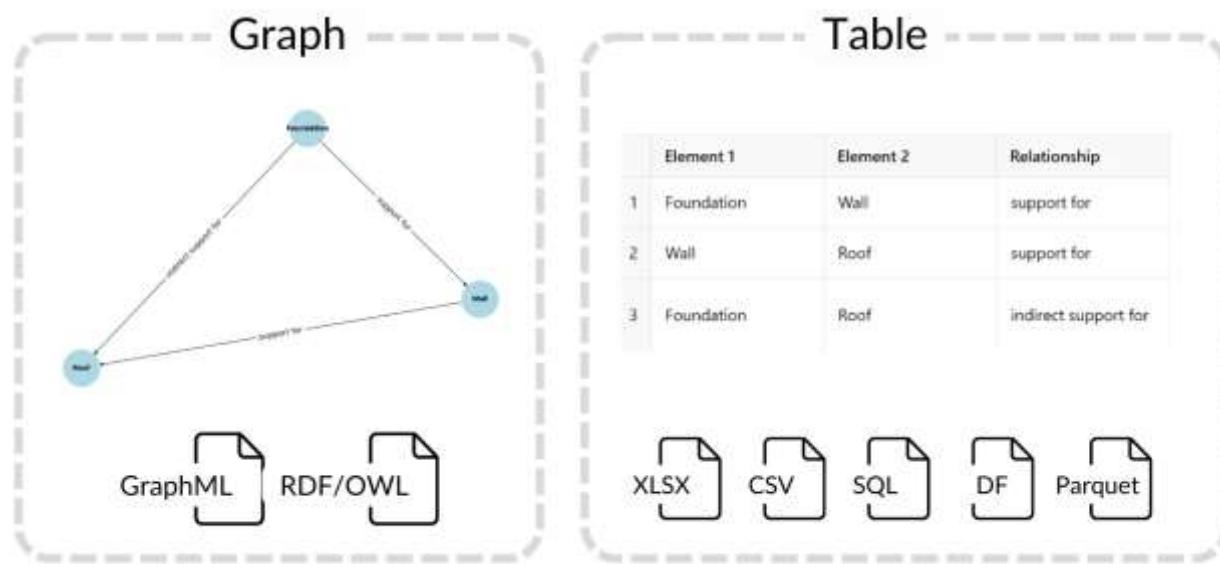
Obr. 6.1-7 Datový model RDF: Uzly, hrany a trojice znázorňující vztahy mezi stavebními prvky.

Teoreticky umožňuje logická inference risonerů (software pro automatickou logickou inferenci) odvozovat z ontologií nové výroky. Například pokud ontologie budovy zaznamenává, že "základ je oporou pro zed" a "zed je oporou pro střechu" (obr. 6.1-7), risoner je schopen automaticky odvodit, že "základ je oporou pro střechu".

Takový mechanismus je užitečný pro optimalizaci analýzy dat, protože se vyhýbá explicitnímu předepisování všech závislostí. Nevytváří však nové znalosti, ale pouze identifikuje a strukturuje již známé skutečnosti.

Sémantika sama o sobě nevytváří nový význam nebo znalost a není v tomto ohledu nadřazena jiným technologiím ukládání a zpracování dat. Reprezentace dat z relačních databází jako trojic je nečiní smysluplnějšími. Nahrazení tabulek grafovými strukturami může být užitečné pro sjednocení datových modelů, snadné vyhledávání a bezpečné úpravy, ale nečiní data "chytrějšími" - počítač nezačne lépe rozumět jejich obsahu.

Logické vztahy v datech lze uspořádat bez složitých sémantických technologií (obr. 6.1-8). Tradiční relační databáze (SQL) i formáty CSV nebo XLSX umožňují budovat podobné závislosti. Například ve sloupcové databázi můžete přidat pole "podpora střechy" a při vytváření stěny automaticky spojit střechu se základem. Tento přístup je implementován bez použití RDF, OWL, grafů nebo risonerů, zůstává jednoduchým a efektivním řešením pro ukládání a analýzu dat.



Obr. 6.1-8 Srovnání datových modelů grafů a tabulek pro reprezentaci stejných logických vztahů.

Rozhodnutí řady velkých stavebních společností a organizace pro vývoj formátu IFC [94] následovat koncept sémantického webu, který se koncem devadesátých let zdál být slibný, mělo významný dopad na vývoj norem ve stavebnictví.

Paradoxem však je, že samotná koncepce sémantického webu, původně určená pro internet, nebyla široce přijata ani ve svém původním prostředí. Navzdory vývoji RDF a OWL se plnohodnotný sémantický web ve své původní koncepci neobjevil a jeho vznik je již nepravděpodobný.

Proč sémantické technologie ve stavebnictví nesplňují očekávání?

Jiná odvětví se potýkají s omezeními technologií pro využití sémantiky. V herním průmyslu se pokusy popsat herní objekty a jejich interakce pomocí ontologií ukázaly jako neúčinné kvůli vysoké dynamice změn. V důsledku toho byly upřednostněny jednodušší datové formáty, jako jsou XML a JSON, spolu s algoritmickými řešeními. Podobná situace byla v odvětví nemovitostí: vzhledem k regionálním rozdílům v terminologii a častým změnám na trhu se použití ontologií ukázalo jako příliš složité, zatímco jednoduché databáze a standardy, jako je RETS [119], se s výzwami výměny dat vyrovnaly lépe.

Technické potíže, jako je složitost značení, vysoká pracnost podpory a nízká motivace vývojářů, zpomalily zavádění sémantického webu i v jiných odvětvích ekonomiky. RDF (Resource Description Framework) se nestal masovým standardem a ontologie se ukázaly jako příliš složité a ekonomicky neopodstatněné.

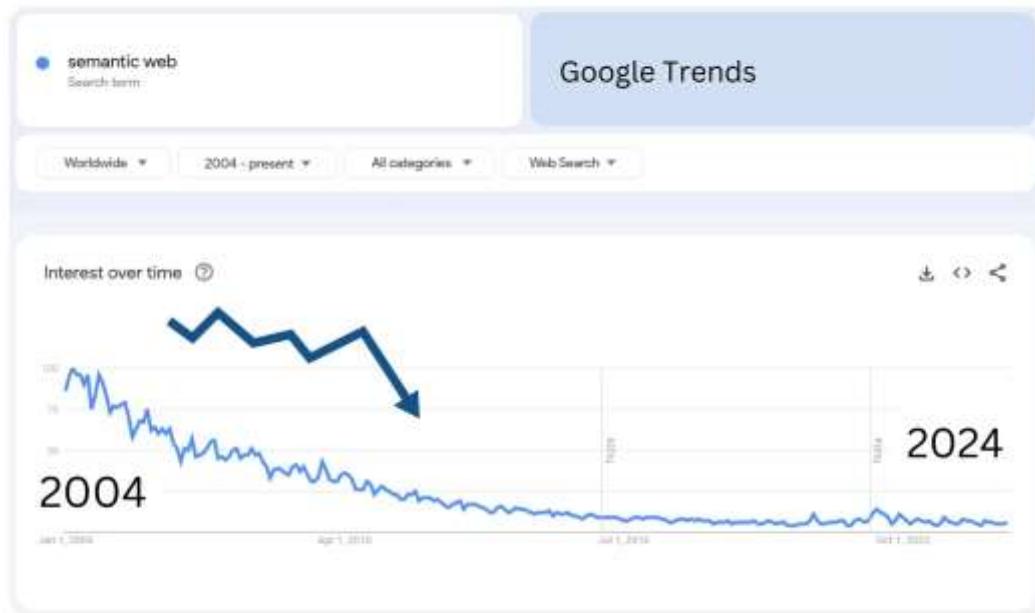
V důsledku toho se nepodařilo uskutečnit ambiciozní myšlenku vytvoření globálního sémantického webu. Přestože některé prvky této technologie, jako jsou ontologie a SPARQL, si našly cestu do podnikových řešení, původního cíle vytvořit jednotnou komplexní datovou strukturu nebylo dosaženo.

Koncepce internetu, v němž jsou počítače schopny dávat obsahu smysl, se ukázala jako technicky obtížná a komerčně nevýhodná. Proto společnosti, které tuto myšlenku podporovaly, nakonec omezily její využití na jednotlivé užitečné nástroje a ponechaly RDF a OWL spíše pro vysoce specializované firemní potřeby než pro internet jako celek. Analýza trendů Google (obrázek 6.1-9) za posledních 20 let naznačuje, že sémantický web už možná nemá žádné vyhlídky.

Není třeba zbytečně množit entity. Pokud existuje několik logicky konzistentních vysvětlení nějakého jevu, která jej vysvětlují stejně dobře, měli bychom při zachování všech ostatních podmínek dát přednost nejjednoduššímu z nich.

břitva

Vystává zde logická otázka: proč vůbec používat při konstrukci tripletů, riserů a SPARQL, když lze data zpracovávat pomocí populárních strukturovaných dotazů (SQL, Pandas, Apache®)? V podnikových aplikacích je SQL standardem pro práci s databázemi. SPARQL naopak vyžaduje složité grafové struktury a specializovaný software a podle trendů v Googlu nepřitahuje zájem vývojářů.



Obr. 6.1-9 Zájem o dotazy "sémantického internetu" podle statistik Google.

Grafové databáze a klasifikační stromy mohou být v některých případech užitečné, ale jejich použití není vždy opodstatněné pro většinu každodenních úloh. V důsledku toho má vytváření znalostních grafů a využívání technologií sémantického webu smysl pouze tehdy, je-li třeba sjednotit data z různých zdrojů nebo realizovat složité logické závěry.

Přechod od tabulek ke grafickým datovým modelům zlepšuje vyhledávání a sjednocuje tok informací, ale nezvyšuje smysluplnost dat pro stroje. Otázkou není, zda by se měly sémantické technologie používat, ale kde mají skutečný význam. Než ve své firmě zavedete ontologii, sémantiku a grafové databáze, zjistěte, které firmy již tyto technologie úspěšně používají a kde se jim to nepodařilo.

Navzdory ambiciozním očekáváním se sémantické technologie nikdy nestaly univerzálním řešením pro strukturování dat ve stavebnictví. V praxi tyto technologie nevedly k univerzálnímu řešení, ale pouze přidaly nové složitosti a tyto snahy jsou opakem nenaplněných ambicí konceptu sémantického internetu, kde očekávání daleko předčila realitu.

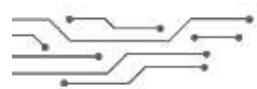


Obr. 6.1-10 Geometrie a informace ve stavebních procesech: od složitých systémů CAD a BIM- po zjednodušená data pro analytiku.

Zatímco v IT byly neúspěchy sémantického webu kompenzovány nástupem nových technologií (big data, IoT, strojové učení, AR/VR), stavebnictví takové příležitosti nemá.

Kromě problémů s používáním konceptů pro sdělování datových vztahů mezi prvky projektu zůstává zásadní problém - samotná dostupnost těchto dat. Ve stavebnictví stále převládají uzavřené systémy, což ztěžuje práci s daty, sdílení informací a zvyšování efektivity procesů.

Právě uzavřenosť dat se stává jednou z hlavních překážek, které brání rozvoji digitálních řešení ve stavebnictví. Na rozdíl od odvětví informačních technologií, kde se standardem staly otevřené a harmonizované datové formáty, v odvětví CAD (BIM) používá každý software svůj vlastní formát, čímž vytváří uzavřené ekosystémy a uměle omezuje uživatele.



KAPITOLA 6.2.

UZAVŘENÉ FORMÁTY PROJEKTŮ A OTÁZKY INTEROPERABILITY

Uzavřená data a klesající produktivita: slepá ulička průmyslu CAD (BIM)

Proprietární charakter systémů CAD -systémů vedl k tomu, že každý program má svůj vlastní jedinečný datový formát, který je buď uzavřený a zvenčí nepřístupný - RVT, PLN, DWG, NDW, NWD, SKP, nebo je k dispozici v polostrukturované podobě prostřednictvím poměrně složitého procesu konverze JSON, XML (CPIXML), IFC, STEP a ifcXML, IfcJSON, BIMJSON, IfcSQL, CSV, atd..

Různé datové formáty, v nichž mohou být uložena stejná data o stejných projektech, se liší nejen strukturou, ale také různými verzemi vnitřního značení, což musí vývojáři zohlednit, aby zajistili kompatibilitu aplikací. Například formát CAD z roku 2025 se otevře v programu CAD z roku 2026, ale stejný projekt se nikdy neotevře ve všech verzích programu CAD, které mohly být k dispozici před rokem 2025.

Tím, že poskytovatel softwaru ve stavebnictví neposkytuje přímý přístup k databázím, často vytváří svůj vlastní jedinečný formát a nástroje, které musí odborník (projektant nebo správce dat) používat pro přístup, import a export dat.

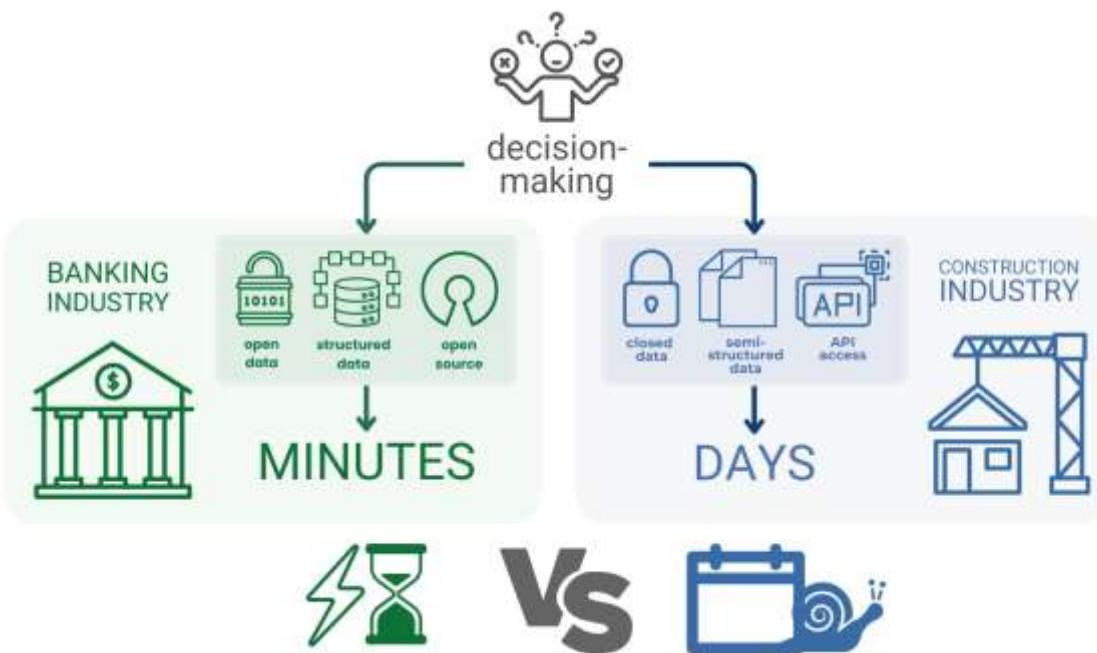
V důsledku toho prodejci základních CAD (BIM) a souvisejících řešení (např. ERP/PMIS)) neustále zvyšují ceny za používání produktů a běžní uživatelé jsou nuceni platit "provizi" v každé fázi přenosu dat podle formátů [63]: za připojení, import, export a práci s daty, která si uživatelé sami vytvořili.

Náklady na přístup k datům v cloudovém úložišti z populárních produktů CAD - (BIM-) dosáhnou v roce 2025 1 dolaru za transakci [120] a předplatné stavebních ERP -produktů pro středně velké firmy dosahuje pěti- a šestimístných částek ročně [121].

Podstatou moderního stavebního softwaru je, že nikoliv automatizace nebo zvýšená efektivita, ale schopnost inženýrů porozumět konkrétnímu vysoce specializovanému softwaru ovlivňuje kvalitu a náklady na zpracování dat stavebních projektů, stejně jako zisky a dlouhodobé přežití firem, které realizují stavební projekty.

Nedostatečný přístup k databázím CAD -systémům, které jsou využívány v desítkách dalších systémů a stovkách procesů [63], a z toho plynoucí nedostatek kvalitní komunikace mezi jednotlivými specialisty přivedl stavebnictví do pozice jednoho z nejfektivnějších odvětví ekonomiky z hlediska produktivity [44].

Za posledních 20 let projektových aplikací CAD- (BIM-), nástupu nových systémů (ERP), nových stavebních technologií a materiálů klesla produktivita celého stavebnictví o 20% (obrázek 2.2-1), zatímco celková produktivita všech odvětví ekonomiky, která nemají zásadní problémy s přístupem k databázím a marketingově podobným konceptům BIM, se zvýšila o 70% (ve zpracovatelském průmyslu o 96%) [122].



Obr. 6.2-1 Vzhledem k izolovanosti a složitosti projektových dat, na nichž jsou ve stavebnictví závislé desítky oddělení a stovky procesů, je rychlosť rozhodování několikanásobně nižší než v jiných odvětvích.

Existují však i ojedinělé příklady alternativních přístupů k vytváření interoperability mezi řešeními CAD. Největší evropská stavební společnost s projektem SCOPE [123], který byl zahájen již v roce 2018, ukazuje, jak je možné překročit klasickou logiku systémů CAD- (BIM-). Namísto snahy podřídit si IFC nebo spoléhání na proprietární geometrická jádra používají vývojáři SCOPE rozhraní API a SDK reverzního inženýrství k získávání dat z různých CAD programů, jejich převodu do neutrálních formátů, jako je OBJ nebo CPIXML, založených na jediném open source geometrickém jádru OCCT, a dále je aplikují na stovky obchodních procesů stavebních a projekčních firem. Navzdory progresivitě myšlenky se však takové projekty potýkají s omezeními a složitostí svobodných geometrických jader a stále součástí uzavřených ekosystémů jedné společnosti, které reprodukují logiku monovenderových řešení.

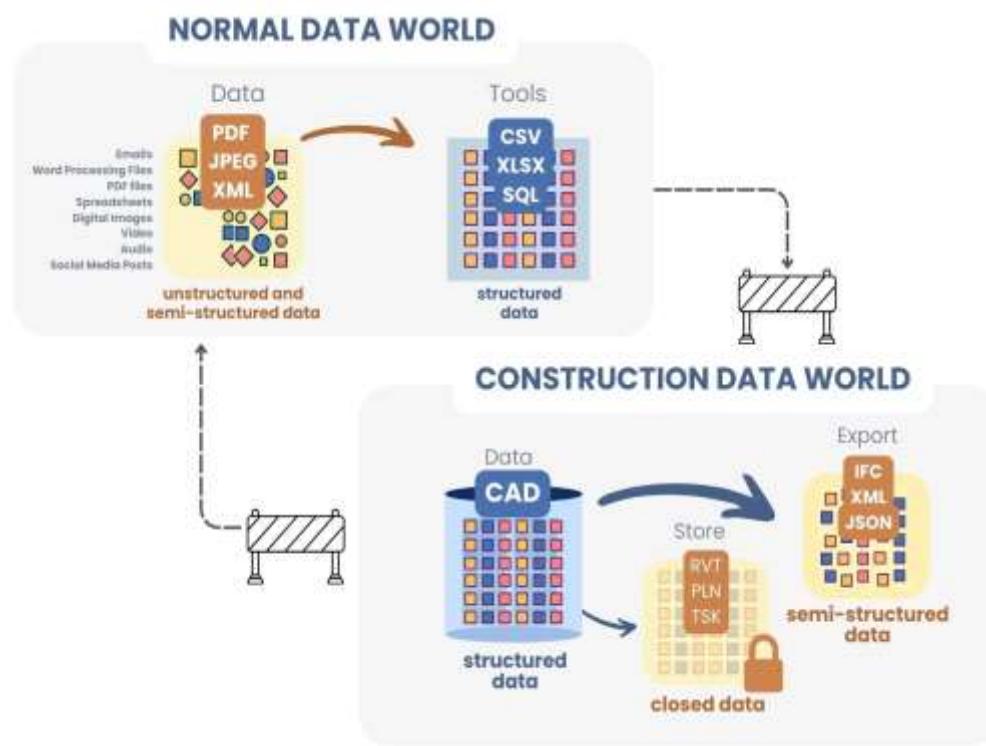
Vzhledem k omezením uzavřených systémů a rozdílům v datových formátech, jakož i nedostatku účinných nástrojů pro jejich sjednocení, se společnosti, které musí pracovat s formáty CAD, potýkají s hromaděním značného množství dat s různou mírou struktury a uzavřenosti. Tato data nejsou rádně využívána a mizí v archivech, kde zůstávají navždy zapomenuta a nevyužita.

Data získaná značným úsilím ve fázi návrhu se stávají pro další použití nedostupná kvůli své složitosti a uzavřenosti.

V důsledku toho byli vývojáři ve stavebnictví v posledních 30 letech nuceni stále znova čelit stejnemu problému: každý nový uzavřený formát nebo proprietární řešení vyvolává potřebu integrace se stávajícími otevřenými a uzavřenými systémy CAD. Tyto neustálé pokusy o zajištění interoperability mezi různými řešeními CAD a BIM pouze komplikují datový ekosystém, místo aby přispívaly k jeho zjednodušení a standardizaci.

Mýtus o interoperabilitě mezi systémy CAD

Jestliže v polovině 90. let bylo hlavním směrem vývoje interoperability v prostředí CAD prolomení proprietárního formátu DWG - což vyvrcholilo vítězstvím aliance Open DWG [75] a skutečným otevřením nejpopulárnějšího výkresového formátu pro celý stavební průmysl -, pak v polovině 20. let se těžiště přesunulo jinam. Ve stavebnictví se prosazuje nový trend: četné vývojové týmy se zaměřují na vytváření takzvaných "mostů" mezi uzavřenými systémy CAD (uzavřený BIM), formátem IFC a otevřenými řešeními (otevřený BIM). Většina těchto iniciativ je založena na využití formátu IFC a geometrického jádra OCCT, které poskytují technický most mezi různými platformami. Tento přístup je považován za slibný směr, který může výrazně zlepšit výměnu dat a interoperabilitu softwarových nástrojů.

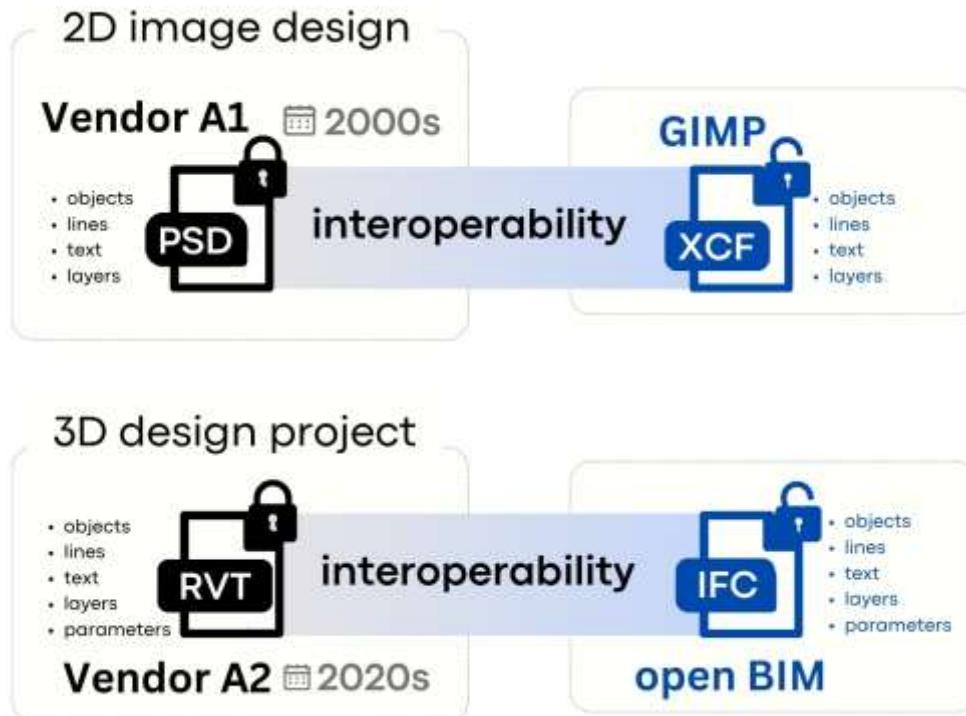


Obrázek 6.2-2 Zatímco ostatní odvětví pracují s otevřenými daty, stavebnictví musí pracovat s uzavřenými nebo volně strukturovanými formáty CAD (BIM).

Tento přístup má historické paralely. V roce 2000 se vývojáři ve snaze překonat dominanci největšího dodavatele grafických editorů (2D world) pokusili vytvořit bezproblémovou integraci mezi jeho proprietárním řešením a svobodným open source - alternativu ke GIMPu (obr. 6.2-3). Tehdy, stejně jako dnes v konstrukci, šlo o snahu propojit uzavřené a otevřené systémy při zachování komplexních parametrů, vrstev a vnitřní logiky softwaru.

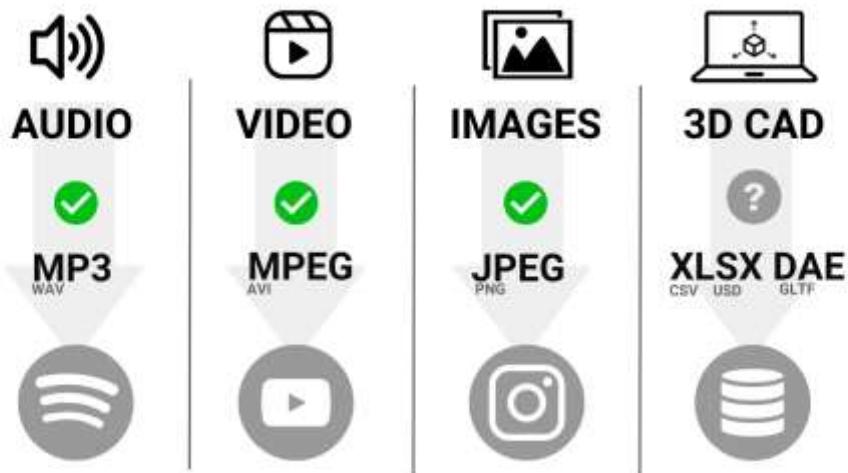
Uživatelé však ve skutečnosti hledali jednoduchá řešení - plochá, otevřená data bez přílišné složitosti vrstev a parametrů programu (obdoba geometrického jádra v CAD). Uživatelé hledali jednoduché a otevřené datové formáty bez nadmerné logiky. V grafice se takovými formáty staly JPEG, PNG a GIF. Dnes se používají v sociálních sítích, na webových stránkách, v aplikacích - jsou snadno zpracovatelné

a interpretovatelné bez ohledu na platformu nebo výrobce softwaru.



Obrázek 6.2-3 Interoperabilita datových formátů ve stavebnictví je podobná cestě od pokusů o sloučení proprietárního produktu populárního výrobce a Open Source GIMP v roce 2000.

V důsledku toho dnes téměř nikdo v odvětví zpracování obrazu nepoužívá uzavřené formáty, jako je PSD nebo otevřený XCF, pro aplikace, sociální sítě, jako je Facebook a Instagram, nebo jako obsah na webových stránkách. Místo toho se pro většinu úloh používají ploché a otevřené formáty JPEG, PNG a GIF kvůli snadnému použití a široké kompatibilitě. Otevřené formáty, jako jsou JPEG a PNG, se staly standardem pro sdílení obrázků díky své univerzálnosti a široké podpoře, což usnadňuje jejich použití na různých platformách. Podobný přechod lze pozorovat i u dalších výměnných formátů, jako je video a audio, kde se zdůrazňují univerzální formáty jako MPEG a MP3 pro jejich účinnost komprese a širokou kompatibilitu. Takový přechod ke standardizaci zjednodušil sdílení a přehrávání obsahu a informací a zpřístupnil je všem uživatelům na různých platformách (obrázek 6.2-4).



Obrázek 6.2-4 Zjednodušené formáty bez složitých editačních funkcí se staly oblíbenými pro sdílení a používání dat.

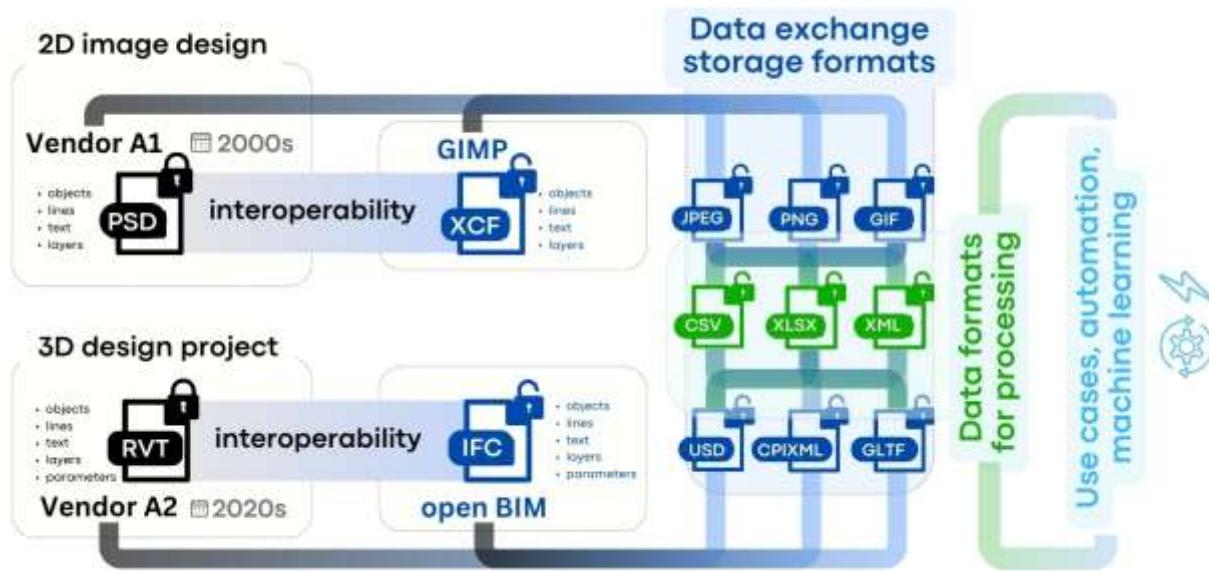
Podobné procesy probíhají i v 3D modelování. Jednoduché a otevřené formáty jako USD, OBJ, glTF, DAE, DXF, SQL a XLSX se v projektech stále více používají pro výměnu dat mimo prostředí CAD (BIM). Tyto formáty uchovávají všechny potřebné informace včetně geometrie a metadat, aniž by bylo nutné pracovat se složitou strukturou BREP, geometrickými jádry nebo interními klasifikátory specifickými pro dodavatele. Proprietární formáty jako NWC, SVF, SVF2, CPIXML a CP2 poskytované předními dodavateli softwaru rovněž plní podobné funkce, ale na rozdíl od otevřených standardů zůstávají uzavřené.

Je pozoruhodné (a stojí za to znova připomenout, jak již bylo zmíněno v předchozí kapitole), že tuto myšlenku - odmítnutí mezilehlých neutrálních a parametrických formátů, jako jsou IGES, STEP a IFC - podpořil již v roce 2000 hlavní dodavatel CAD, který vytvořil BIM Whitepaper a v roce 1994 zaregistroval formát IFC. V dokumentu Whitepaper 2000 "Integrated Design and Manufacturing" (Integrovaný návrh a výroba). [65] dodavatel CAD zdůrazňuje důležitost nativního přístupu k databázi CAD v rámci softwarového prostředí bez nutnosti používat zprostředkující překladače a parametrické formáty, aby byla zachována úplnost a přesnost informací.

Stavební průmysl se dosud nedohodl ani na nástrojích pro přístup k databázím CAD, ani na jejich nuceném zpětném inženýrství, ani na přijetí společného zjednodušeného formátu dat pro použití mimo platformy CAD (BIM). Například mnoho velkých společností ve střední Evropě a v německy mluvících oblastech působících ve stavebnictví používá ve svých ERP -systémech formát CPIXML [121]. Tento proprietární formát, který je druhem XML, spojuje data CAD (BIM) projektu, včetně geometrických a metadat, do jedné organizované zjednodušené struktury. Velké stavební společnosti také vytvářejí nové vlastní formáty a systémy, jako je tomu v případě projektu SCOPE, o kterém jsme hovořili v předchozí kapitole

Uzavřená logika parametrických formátů CAD nebo komplexních parametrických souborů IFC (STEP)

je ve většině obchodních procesů nadbytečná. Uživatelé hledají zjednodušené a ploché formáty, jako jsou USD, CPIXML, XML &OBJ, DXF, glTF, SQLite, DAE a další.XLSX, které obsahují všechny potřebné informace o prvcích, ale nejsou zatíženy nadbytečnou geometrickou logikou BREP, závislostí na geometrických jádrech a vnitřními klasifikacemi konkrétních CAD a BIM -produktů (obr. 1). 6.2-5).



Obr. 6.2-5 Pro většinu případů použití volí uživatelé nejjednodušší možné formáty, které jsou nezávislé na programech dodavatele.

Nástup plochých obrazových formátů, jako jsou JPEG, PNG a GIF, zbavených nadbytečné logiky interních motorů výrobců, podpořil vývoj tisíců interoperabilních řešení pro zpracování a využití grafiky. To vedlo ke vzniku aplikací od retušovacích a filtračních nástrojů až po sociální sítě, jako jsou Instagram, Snapchat a Canva, kde lze tato zjednodušená data využívat bez vazby na konkrétního výrobce softwaru.

Standardizace a zjednodušení návrhových CAD -formátů podnítí vznik mnoha nových uživatelsky přívětivých a nezávislých nástrojů pro práci se stavebními projekty.

Odklon od složité logiky aplikací dodavatelů vázaných na uzavřená geometrická jádra k univerzálním otevřeným formátům založeným na knihovnách zjednodušených prvků vytváří předpoklady pro flexibilnější, transparentnější a efektivnější zpracování dat. Tím se také otevírá přístup k informacím pro všechny strany zapojené do procesu výstavby - od projektantů po zákazníky a servisní služby.

Přesto je velmi pravděpodobné, že se v příštích letech dodavatelé CAD pokusí opět posunout debatu o interoperabilitě a přístupu k databázím CAD. Půjde již o "nové" koncepty - jako jsou granulární data, inteligentní grafy, "federované modely", digitální dvojčata v cloudových úložištích - a také o vytváření oborových aliancí a standardů, které budou pokračovat v cestě BIM a otevřeného BIM. Navzdory atraktivní terminologii se tyto iniciativy mohou opět stát nástroji k udržení uživatelů v rámci proprietárních ekosystémů. Jedním z příkladů je aktivní propagace formátu USD (Universal Scene

Description) jako "nového standardu" pro spolupráci mezi platformami CAD (BIM) od roku 2023.

Přejít na a granulární data m

Vznik aliance AOUSD [124] v roce 2023 znamená důležitý obrat ve stavebnictví. Jsme svědky počátku nové reality, kterou utvářejí dodavatelé CAD, při nakládání se stavebními daty prostřednictvím několika významných změn. První významná změna se týká vnímání dat CAD -data. Odborníci, kteří se podílejí na raných fázích koncepčního návrhu, si stále více uvědomují, že vytvoření návrhu v prostředí CAD je pouze výchozím bodem. Data vytvořená během procesu navrhování se nakonec stávají základem pro analýzu, provoz a správu objektů. To znamená, že musí být přístupná a použitelná v systémech nad rámec tradičních nástrojů CAD.

Současně dochází k revoluci v přístupu předních vývojářů. Přední dodavatel CAD-, tvůrce konceptu BIM a formátu IFC, provádí ve své strategii nečekaný obrat. Od roku 2023 upouští od tradičního ukládání dat v samostatných souborech, zaměřuje se na práci s granulovanými (normalizovanými a strukturovanými) daty a přechází na datově orientovaný přístup [125].

Prodejci sledují historické trendy jiných odvětví: většina uživatelů nepotřebuje uzavřené formáty CAD (podobné PSD) nebo složité parametrické soubory IFC (podobné GIMPu s logikou vrstev). Potřebují jednoduché obrázky objektů, které lze použít v CAFM (stavební Instagram), ERP (Facebook) a tisíce dalších procesů plných tabulek Excel a dokumentů PDF.

Současné trendy ve stavebnictví potenciálně vytvářejí podmínky pro postupný odklon od parametrických a složitých formátů ve prospěch univerzálnějších a nezávislejších formátů USD, GLTF, DAE, OBJ (s metainformacemi jak v rámci hybridních, tak v samostatných strukturovaných nebo volně strukturovaných formátech). Historičtí lídři, včetně velkých projekčních firem, které kdysi v polovině 90. let aktivně prosazovaly IFC, nyní otevřeně propagují nový formát USD [93] a zdůrazňují jeho jednoduchost a univerzálnost (obr. 6.2-6). Masové přijetí USD v produktech, kompatibilita s GLTF a aktivní integrace do nástrojů, jako jsou Blender, Unreal Engine a Omniverse, ukazují potenciál počátku nového paradigmatu práce s daty. Spolu s popularitou lokalizovaných řešení, jako je evropský plochý formát USD - CPIXML, používaný v populárních evropských ERP, by mohl USD potenciálně posílit pozici ve střední Evropě. Organizace podílející se na vývoji formátu IFC již přizpůsobují svou strategii USD [126], což jen potvrzuje nevyhnutelnost tohoto posunu.

Technical Specifications				Comparison / Notes
File Structure	Monolithic file	Uses ECS and linked data	IFC stores all data in one file; USD uses Entity-Component-System and linked data for modularity and flexibility	
Data Structure	Complex semantics, parametric geometry	Flat format, geometry in MESH, data in JSON	IFC is complex and parametric; USD is simpler and uses flat data	
Geometry	Parametric, dependent on BREP	Flat, MESH (triangular meshes)	IFC uses parametrics; USD uses meshes for simplified processing.	
Properties	Complex structure of semantic descriptions	Properties in JSON, easy access	Properties in USD are easier to use thanks to JSON	
Export/Import	Complex implementation, dependent on third-party SDKs	Easy integration, wide support	USD integrates more easily and is supported in many products	
Format Complexity	High, requires deep understanding	Low, optimized for convenience	The time required to understand the structure of the file and the information stored in it.	
Performance	Can be slow when processing large models	High performance in visualization and processing	USD is optimized for speed and efficiency. Simulations, machine learning, AI, smart cities will be held in the Nvidia Omniverse	
Integration with 3D Engines	Limited	High, designed for graphics engines	USD excels with native support for real-time visualization platforms	
Support outside CAD Software	BlenderBIM, IfcOpenShell	Unreal Engine, Unity, Blender, Omniverse	USD is widely supported in graphics tools	
Cloud Technology Support	Limited	Well-suited for cloud services and online collaboration	USD is optimized for cloud solutions	
Ease of Integration into Web Applications	Difficult to integrate due to size and complexity	Easy to integrate, supports modern web technologies	USD is preferable for web applications	
Change Management	Versions through separate files	Versioning built into the format core	IFC handles changes via separate files, while USD embeds versioning directly into its structure	
Collaboration Support	Supports data exchange between project participants	Designed for collaborative work on complex scenes	USD provides efficient collaboration through layers and variations	
Learnability	Steep learning curve due to complexity	Easier to master thanks to a clear structure	USD is easier to learn and implement	

Obrázek 6.2-6 Srovnání technických specifikací formátů IFC a USD.

Na tomto pozadí má USD potenciál stát se de facto standardem a slibuje překonat mnohá současná omezení, která souvisejí především se složitostí stávajících formátů CAD - (BIM-) a závislostí jejich interpretace na geometrických jádřech.

Namísto parametrických a složitých formátů CAD a IFC - získají ve stavebnictví místo zjednodušené datové formáty USD, glTF, DAE, OBJ s metainformacemi o prvcích ve formátu CSV, XLSX, JSON, XML díky své jednoduchosti a flexibilitě.

Současné změny ve stavebnictví na první pohled vypadají jako technologický průlom spojený s přechodem od zastaralé IFC k modernějšímu USD. Je však třeba vzít v úvahu, že již v roce 2000 stejný

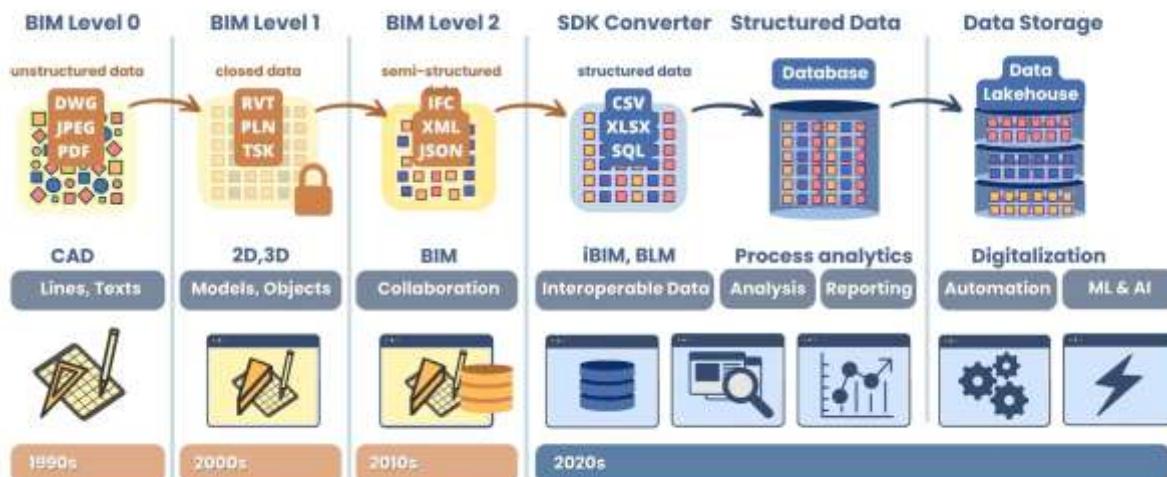
dodavatel CAD, který vyvinul IFC, psal o jeho problémech a potřebě přístupu k databázi [65] a nyní aktivně propaguje přechod na nový standard - USD.

Za další fasádou "otevřených dat" USD a "nových" konceptů granulární správy dat prostřednictvím cloudových aplikací, které začínají prosazovat dodavatelé CAD, se může skrývat záměr dodavatelů monopolizovat správu projektových dat, kdy se uživatelé ocitají v situaci, kdy volba formátu souvisí spíše s firemními zájmy než se skutečnými potřebami.

Analýza klíčových skutečností [93] ukazuje, že hlavním účelem těchto změn není ani tak uživatelský komfort, jako především udržení kontroly nad ekosystémy a datovými toky ve prospěch dodavatelů, kteří za 40 let nikdy nebyli schopni poskytnout přístup k databázím CAD.

Možná nastal čas, aby společnosti přestaly čekat na nové koncepty od dodavatelů softwaru a zaměřily se na vlastní vývoj v datově orientovaném směru. Poté, co se zbaví problémů s přístupem k datům prostřednictvím nástrojů reverzního inženýrství, bude moci odvětví samostatně přejít na moderní, bezplatné a pohodlné nástroje pro práci a analýzu dat bez vnukování nových konceptů.

CAD (BIM) Maturity Levels: From Stage 0 to Structured Data



Obr. 6.2-7 Úroveň vyspělosti CAD (BIM): od nestrukturovaných dat ke strukturovaným datům a úložištěm.

Přístup k databázím, otevřeným datům a formátům se nevyhnutelně stane ve stavebnictví standardem, bez ohledu na snahu dodavatelů tento proces zdržovat - je to jen otázka času (obr. 6.2-7). Tempo tohoto přechodu se může výrazně zvýšit, pokud se stále více odborníků seznámí s otevřenými formáty, databázovými nástroji a dostupnými SDK pro reverzní inženýrství, které umožňují přímý přístup k datům CAD -systémům [92].

Budoucnost je v otevřených, jednotných a analyticky přístupných datech. Aby se stavební a inženýrské společnosti vyhnuly závislosti na řešeních dodavatelů a nestaly se rukojmími uzavřených ekosystémů, budou muset dříve či později vsadit na otevřenosť a nezávislost a zvolit formáty a řešení, která poskytují plnou kontrolu nad daty.

Data, která dnes vznikají ve stavebnictví, budou v budoucnu klíčovým zdrojem pro obchodní rozhodnutí. Budou fungovat jako strategické "palivo", které pohání rozvoj a efektivitu stavebních společností. Budoucnost stavebnictví spočívá ve schopnosti pracovat s daty, nikoli ve volbě datových formátů nebo modelů.

Pro pochopení rozdílu mezi otevřenými formáty USD, glTF, DAE, OBJ a proprietárními parametrickými formáty CAD je důležité vzít v úvahu jeden z nejsložitějších a klíčových datových prvků vizualizace a konstrukčních výpočtů - geometrii a procesy jejího generování. A abychom pochopili, jak se geometrická data stávají základem analytických a výpočtů ve stavebnictví, je třeba se hlouběji podívat na mechanismy generování, transformace a ukládání geometrie.



KAPITOLA 6.3.

GEOMETRIE VE STAVEBNICTVÍ: OD ČAR PO METRY KRYCHLOVÉ

Když se čáry mění v peníze aneb proč stavitele potřebují geometrii

Geometrie ve stavebnictví není jen vizualizace, ale také základ pro přesné kvantitativní výpočty. V modelu projektu geometrie doplňuje seznamy parametrů prvků (obr. 3.1-16) o důležité objemové charakteristiky, jako jsou délka, plocha a objem. Tyto hodnoty objemových parametrů se vypočítávají automaticky pomocí geometrických jader a jsou výchozím bodem pro odhad, harmonogramy a modely zdrojů. Jak jsme již uvedli v 5. části této knihy a v kapitole "Kalkulace a odhad stavebních projektů", jsou to právě objemové parametry skupin objektů z CAD -modelů, které tvoří základ moderních ERP, PMIS -systémů/ Geometrie hraje zásadní roli nejen ve fázi projektování, ale také při řízení realizace projektu, kontrole harmonogramu, rozpočtování a provozu. Stejně jako před tisíci lety při stavbě egyptských pyramid závisela přesnost projektu na délkových měřích, jako jsou lokty a lokte, dnes přesnost interpretace geometrie v CAD -programech přímo ovlivňuje výsledek: od rozpočtu a termínů až po výběr dodavatele a logistiku dodávky

Ve vysoce konkurenčním prostředí s omezeným rozpočtem se přesnost objemových výpočtů, která přímo závisí na geometrii, stává faktorem přežití. Moderní ERP -systémy jsou přímo závislé na správných objemových charakteristikách získaných z CAD - a BIM - modelů. Proto přesný geometrický popis prvků není pouhou vizualizací, ale klíčovým nástrojem pro řízení nákladů a času výstavby.

Historicky byla geometrie hlavním jazykem technické komunikace. Od čar na papyru až po digitální modely sloužily výkresy a geometrická zobrazení jako prostředek výměny informací mezi projektanty, supervizory a odhadci. Před nástupem počítačů se výpočty prováděly ručně pomocí pravítek a úhloměrů. Dnes je tento úkol automatizován díky objemovému modelování: geometrická jádra softwaru CAD převádějí čáry a body na trojrozměrná tělesa, z nichž jsou automaticky extrahovány všechny potřebné charakteristiky.

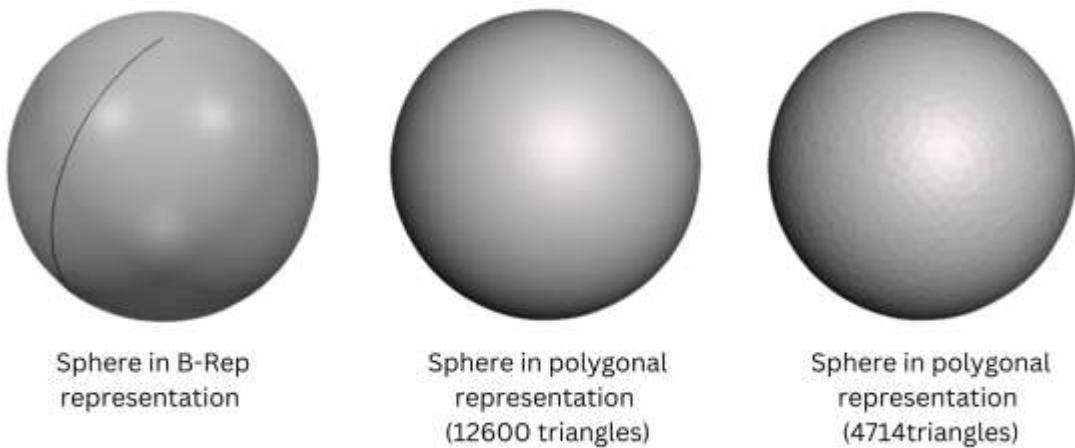
Při práci v programech CAD se vytváření geometrických prvků pro výpočty provádí prostřednictvím uživatelského rozhraní programů CAD- (BIM-). K transformaci bodů a linií na objemová tělesa se používá geometrické jádro, které plní klíčovou úlohu - transformaci geometrie na objemové modely, z nichž se po approximaci automaticky vypočítají objemové charakteristiky prvku.

Od čar k objemům: Jak se z plochy a objemu stávají data

V inženýrské praxi se objemy a plochy počítají z geometrických ploch popsaných analyticky nebo pomocí parametrických modelů, jako je NURBS (nonuniform rational B-splines) v rámci BREP (boundary element representation).

NURBS (Non-Uniform Rational B-Splines) je matematický způsob popisu křivek a ploch, zatímco BREP je rámec pro popis kompletní trojrozměrné geometrie objektu, včetně jeho hranic, které lze definovat pomocí NURBS.

Navzdory přesnosti metod BREP a NURBS vyžadují výkonné výpočetní zdroje a složité algoritmy. Přímý výpočet z takto matematicky přesných popisů je však často výpočetně náročný, proto se v praxi téměř vždy používá tessellace - transformace ploch na síť trojúhelníků -, která zjednoduší následné výpočty. Tesselace je rozdělení složitého povrchu na trojúhelníky nebo mnohoúhelníky. V prostředích CAD /CAE se tato metoda používá pro vizualizaci, objemové výpočty, vyhledávání kolizí, export do formátů jako MESH a analýzu kolizí. Příkladem z přírody je včelí plástev, kde se složitý tvar rozkládá na pravidelnou síť (obr. 6.3-1).



Obr. 6.3-1 Stejná koule v parametrickém popisu BREP a polygonálním zobrazení různým počtem trojúhelníků.

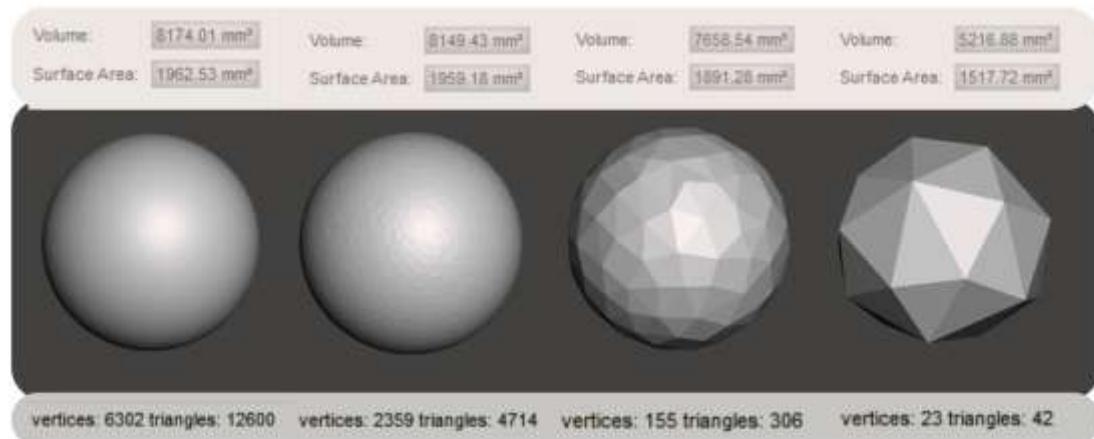
BREP (NURBS), používaný v CAD, není základním modelem geometrie. Byl vytvořen jako vhodný nástroj pro reprezentaci kružnic a racionálních splajnů a pro minimalizaci ukládání geometrických dat. Má však svá omezení - například neschopnost přesně popsat sinusoidu, která je základem šroubovic a ploch, a nutnost používat složitá geometrická jádra.

Naproti tomu trojúhelníkové sítě a tessellace parametrických tvarů se vyznačují jednoduchostí, efektivním využitím paměti a schopností zpracovat velké množství dat (obr. 6.3-2). Tyto výhody umožňují obejít se při výpočtu geometrických tvarů bez složitých a drahých geometrických jader, a v nich obsažených desítek milionů řádků kódu.

Ve většině případů budov nezáleží na tom, jak přesně jsou objemové charakteristiky definovány - prostřednictvím parametrických modelů (BREP, IFC) nebo prostřednictvím polygonů (USD, glTF, DAE, OBJ). Geometrie zůstává formou aproximace: ať už prostřednictvím NURBS nebo MESH, vždy se jedná o přibližný popis tvaru.

Geometrie definovaná jako polygony nebo BREP (NURBS) zůstává do jisté míry pouze způsobem přibližného popisu spojitého tvaru. Stejně jako Fresnelovy integrály nemají přesné analytické vyjádření, je diskretizace geometrie pomocí polygonů nebo NURBS vždy approximací, stejně jako trojúhelníková MESH.

Parametrická geometrie ve formátu BREP je nutná především tam, kde je důležitá minimální velikost dat a kde je možné pro její zpracování a zobrazení použít na zdroje náročná a drahá geometrická jádra. Nejčastěji je to charakteristické pro vývojáře CAD -programů, kteří za tímto účelem používají ve svých produktech geometrická jádra MCAD -dodavatelů. V tomto případě se i v rámci těchto programů BREP -modely v procesu tessellace pro vizualizaci a výpočty často převádějí na trojúhelníky (podobně jako se zjednoduší PSD-soubory do JPEG).



Obr. 6.3-2 Rozdíl objemových charakteristik u obrazců s různým počtem mnohoúhelníků.

Polygonální MESH, stejně jako parametrický BREP, mají své vlastní výhody a omezení, ale cíl je stejný - popsat geometrii s ohledem na úkol uživatele. Přesnost geometrického modelu nakonec závisí nejen na způsobu jeho reprezentace, ale také na požadavcích konkrétní úlohy.

Ve většině konstrukčních problémů může být potřeba parametrické geometrie a složitých geometrických jader nadbytečná.

V každé konkrétní úloze automatizace výpočtu stojí za zvážení, zda význam parametrické geometrie není přehnaný ze strany vývojářů CAD, kteří mají zájem na propagaci a prodeji vlastních softwarových produktů.

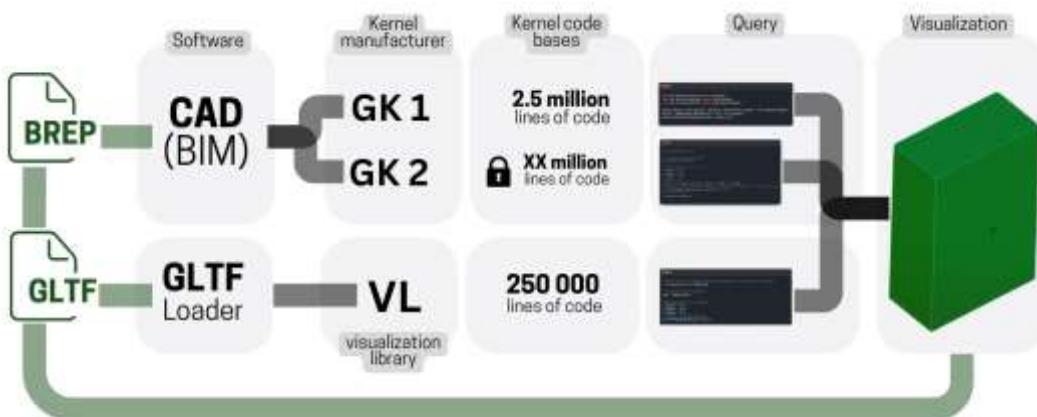
Přechod na MESH, USD a polygony: použití teselace pro geometrii

Ve stavebnictví je při streamování, vývoji systémů, databází nebo automatizaci procesů, které pracují s konstrukčními informacemi a geometrií prvků, důležité usilovat o nezávislost na konkrétních CAD editorech a geometrických jádrech.

Výměnný formát, který se bude používat jak ve výpočetních odděleních, tak na staveništi, by neměl být založen na konkrétním programu CAD- (BIM-). Geometrické informace by měly být ve formátu reprezentovány přímo prostřednictvím tesselace, bez odkazu na geometrické jádro nebo architekturu CAD.

Parametrickou geometrii z CAD lze považovat za mezistupeň, nikoli však za základ univerzálního formátu. Většina parametrických popisů (včetně BREP a NURBS) je v každém případě pro další zpracování převedena na polygonální MESH. Pokud je výsledek stejný (tesselace a polygony) a proces jednodušší, je volba zřejmá. Je to podobné jako volba mezi grafovými ontologiemi a strukturovanými tabulkami (o nichž jsme hovořili ve čtvrté části): přílišná složitost je málokdy opodstatněná (obr. 3.2-10, obr. 6.1-8).

Otevřené formáty, jako jsou: OBJ, STL, glTF, SVF, CPIXML, USD a DAE, používají univerzální strukturu trojúhelníkové sítě, což jim dává značné výhody. Tyto formáty mají vynikající interoperabilitu - lze je snadno číst a vizualizovat pomocí dostupných knihoven s otevřeným zdrojovým kódem, aniž by bylo nutné používat složitá specializovaná geometrická jádra obsahující miliony řádků kódu (obrázek 6.3-3). Tyto univerzální geometrické formáty se používají v aplikacích od relativně jednoduchých nástrojů pro navrhování kuchyní v IKEA™ až po komplexní systémy vizualizace objektů v kinech a VR-aplikacích. Důležitou výhodou je dostupnost velkého množství bezplatných a otevřených knihoven pro práci s těmito formáty, které jsou k dispozici pro většinu platform a programovacích jazyků.



Obrázek 6.3-3 Stejně reprezentace geometrie lze dosáhnout pomocí parametrických formátů a geometrických jader, nebo pomocí triangulačních formátů a vizualizačních knihoven s otevřeným zdrojovým kódem.

Stejně jako samotní uživatelé se i dodavatelé CAD potýkají s problémy při interpretaci zahraničních parametrických CAD formátů nebo otevřených IFC kvůli odlišným geometrickým jádrům. V praxi všichni dodavatelé CAD -prodejci bez výjimky používají k přenosu dat mezi systémy reverzní inženýrský SDK a žádný z nich se pro účely interoperability nespoléhá na formáty jako IFC nebo USD [93].

Namísto používání konceptů propagovaných aliancemi dodavatelů CAD-, které sami nepoužívají, je pro vývojáře a uživatele CAD řešení produktivnější zaměřit se na pochopení výhod každého přístupu v konkrétním kontextu a zvolit si ten či onen typ geometrie v závislosti na případu použití. Volba mezi různými geometrickými reprezentacemi je kompromisem mezi přesností, výpočetní efektivitou a praktickými potřebami konkrétní úlohy.

Složitost spojená s používáním geometrických jader, kterou tradičně vnucují stavebnictví velcí dodavatelé při zpracování dat o návrhu, se často ukazuje jako nadbytečná. Formát USD založený na geometrii MESH se může stát pro průmysl jakousi "Pandořinou skříňkou", která vývojářům otevře nové možnosti organizace výměny dat - mimo rámec struktur IFC a parametrických struktur BREP typických pro dodavatele CAD.

Při bližším pohledu na strukturu formátů USD, DAE, glTF, OBJ atd. je zřejmé, že existují jednodušší, otevřené formáty, které umožňují efektivně organizovat přenos a použití geometrických informací bez nutnosti spoléhat se na složité parametry a uzavřená geometrická jádra. Tento přístup nejen snižuje technický práh vstupu pro vývojáře, ale také podporuje vývoj flexibilních, škálovatelných a skutečně otevřených řešení pro digitální výstavbu.

LOD, LOI, LOMD - jedinečná klasifikace detailů v CAD (BIM).

Kromě formátů geometrické reprezentace ve světě, kde různá odvětví používají různé úrovně podrobnosti a hloubky dat, nabízejí metodiky CAD - (BIM-) své vlastní jedinečné klasifikační systémy, které strukturují přístup k informování o modelech budov.

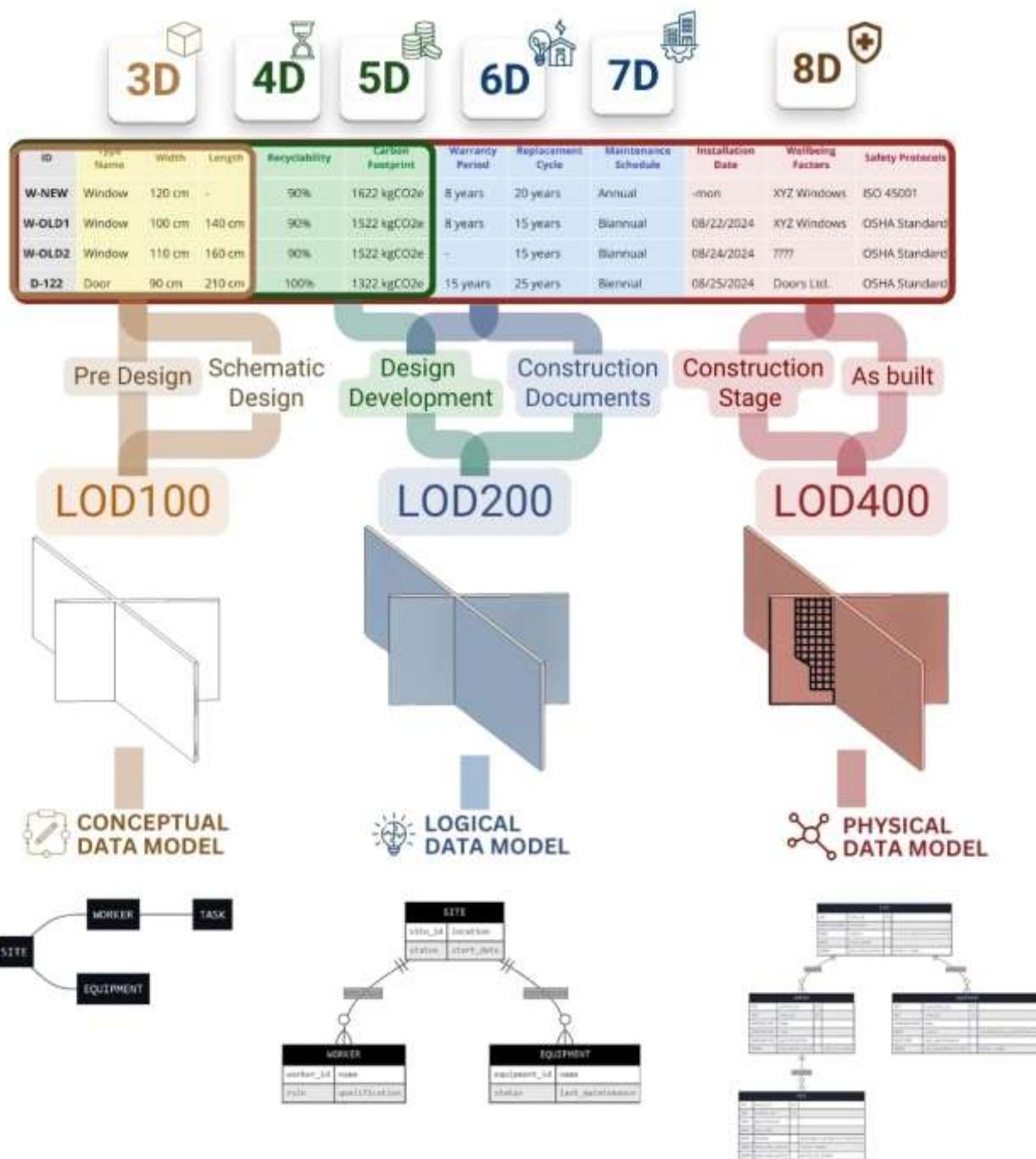
Jedním z příkladů nových přístupů ke standardizaci je zavedení úrovní vývoje modelu, které odrážejí stupeň připravenosti a spolehlivosti grafických i informačních prvků. Pro rozlišení informačního obsahu při práci s daty CAD - (BIM-) se objevily LOD (Level Of Detail) - úroveň podrobnosti grafické části modelu, a LOI (Level Of Information) - úroveň rozpracovanosti dat. Pro integrovaný přístup byl navíc zaveden pojem LOA (Level of Accuracy) - přesnost prezentovaných prvků a LOG (Level of Geometry) pro určení přesnosti grafického zobrazení.

Úrovně podrobnosti (LOD) jsou označeny čísly od 100 do 500, která odrážejí stupeň rozpracovanosti modelu. LOD 100 je koncepční model s obecnými tvary a rozměry. LOD 200 zahrnuje přesnější rozměry a tvary, ale s podmíněnou podrobností. LOD 300 je podrobný model s přesnými rozměry, tvary a umístěním prvků. LOD 400 obsahuje podrobné informace nezbytné pro výrobu a instalaci prvků. LOD 500 odráží skutečný stav zařízení po výstavbě a používá se pro provoz a údržbu. Tyto úrovně popisují strukturu nasycení modelu CAD (BIM) informacemi v různých fázích životního cyklu, včetně 3D, 4D, 5D a dalších.

V reálných projektech je vysoká úroveň podrobnosti (LOD400) často nadbytečná a stačí použít geometrii LOD100 nebo dokonce ploché výkresy, zatímco ostatní data lze získat buď výpočtem, nebo ze souvisejících prvků, které nemusí mít odlišnou geometrii. Například prostory a prvky místností (kategorie prvků místností) nemusí mít žádnou vizuální geometrii, ale obsahují značné množství informací a databází, kolem nichž je postaveno mnoho obchodních procesů.

Proto je důležité před zahájením návrhu jasně definovat požadovanou úroveň podrobnosti. Pro případy použití 4D -7D často stačí i výkresy DWG a minimální geometrie LOD100. Klíčovým úkolem v procesu tvorby požadavků je najít rovnováhu mezi bohatostí a praktičností modelu.

Pokud budeme data CAD (BIM) považovat za databázi (což je), není popis nasycení modelu novými zkratkami v podstatě ničím jiným než postupným modelováním dat pro informační systémy, počínaje konceptuální úrovní a konče fyzickou (obr. 6.3-4), které bylo podrobně rozebráno ve třetí a čtvrté části knihy. Každé zvýšení LOD a LOI znamená přidání informací potřebných pro nové úlohy: výpočty, řízení stavby, provoz a je charakterizováno postupným obohacováním modelu o další informační vrstvy (3D -8D) v podobě různých parametrů, o nichž jsme hovořili v pátém díle knihy.



Obrázek 6.3-4 Proces upřesňování detailů projektu je totožný s datovým modelováním od koncepčního po fyzický datový model.

Geometrie je jen částí návrhových dat, jejichž potřeba není ve stavebních projektech vždy opodstatněná, a klíčovým problémem práce s daty CAD není ani tak způsob vizualizace modelů, ale spíše to, jak lze data z těchto modelů využít mimo programy CAD (BIM -).

V polovině roku 2000 se stavebnictví potýkalo s bezprecedentní výzvou v podobě rychlého nárůstu množství dat v systémech pro správu a zpracování dat, zejména těch, která pocházejí z oddělení CAD

(BIM). Tento dramatický nárůst objemu dat zaskočil manažery firem, kteří nebyli připraveni na rostoucí nároky na kvalitu a správu dat.

Nové standardy CAD (BIM) - AIA, BEP, IDS, LOD, COBie.

S využitím nedostatečného otevřeného přístupu k databázím CAD a omezené konkurence na trhu zpracování dat a s využitím marketingových kampaní spojených s novou zkratkou BIM, začaly organizace zabývající se vývojem přístupů k nakládání s daty CAD - vytvářet nové normy a koncepce, které by měly být zaměřeny na zlepšení postupů správy dat.

Přestože téměř všechny iniciativy přímo či nepřímo podporované dodavateli CAD a vývojáři (BIM) byly zaměřeny na optimalizaci pracovních postupů, vyústily v množství norem, za které lobbovaly různé zúčastněné strany, což vedlo ve stavebnictví k určité nejasnosti a zmatku ohledně datových procesů.

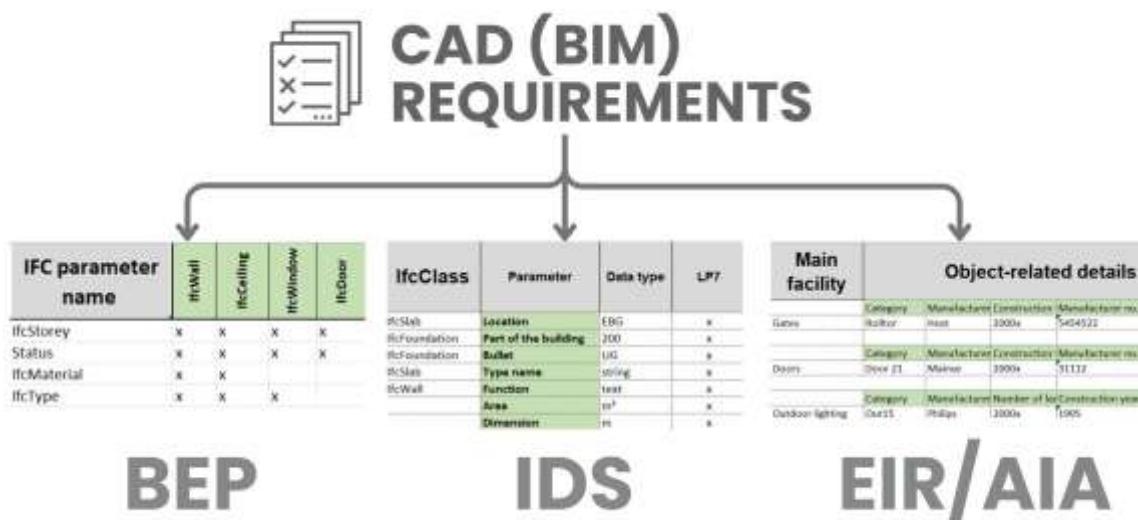
Uveděme si některé nové datové standardy, kromě LOD, LOI, LOA, LOG, které se v posledních letech objevily ve stavebnictví:

- **BEP** (BIM Execution Plan) - popisuje, jak integrovat a využívat CAD (BIM) v projektu, definuje metody a procesy zpracování dat.
- **Dokument EIR /AIA** (Požadavky zadavatele na informace) - připravený zadavatelem před vyhlášením výběrového řízení a obsahuje požadavky na dodavatele, aby připravil a poskytl informace. Slouží jako podklad pro BEP v příslušném projektu.
- **AIM** (Asset Information Model) je součástí procesu BIM. Po předání a dokončení projektu se datový model nazývá Asset Information Model neboli AIM. Účelem AIM je správa, údržba a provoz realizovaného aktiva.
- **IDS** (Information Delivery Specification) - definuje požadavky a to, jaká data a v jakém formátu jsou požadována v různých fázích stavebního projektu.
- **iLOD** je úroveň podrobnosti LOD, s níž jsou informace reprezentovány v BIM -modelu. Definuje, jak podrobné a úplné jsou informace v modelu, od základních geometrických zobrazení až po podrobné specifikace a údaje.
- **eLOD** - LOD úroveň podrobnosti jednotlivých prvků v modelu CAD (BIM). Definuje rozsah, v jakém jsou jednotlivé prvky modelovány, a související informace, jako jsou rozměry, materiály, výkonnostní charakteristiky a další relevantní atributy.
- **APS** (Platform Services) a další produkty hlavních dodavatelů CAD (BIM) - popisují nástroje a infrastrukturu potřebnou k vytváření propojených a otevřených datových modelů.

Ačkoli deklarovaným účelem zavádění standardů CAD (BIM) - jako jsou LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD - je zlepšit kvalitu správy dat a rozšířit možnosti automatizace, v praxi jejich používání často vede k nadmerné složitosti a roztríštěnosti procesů. Pokud budeme považovat CAD (BIM) model za druh databáze, je zřejmé, že mnohé z těchto standardů duplikují dlouhodobě zavedené

a efektivní přístupy používané v jiných odvětvích při práci s informačními systémy. Namísto zjednodušení a sjednocení tyto iniciativy často vytvářejí další terminologickou zátěž a brání zavádění skutečně otevřených a flexibilních řešení.

Je pozoruhodné, že mnohé z těchto nových koncepcí ve skutečnosti nahrazují procesy modelování a ověřování dat, které byly podrobně popsány v prvních částech knihy a které se již dlohu používají v jiných odvětvích ekonomiky. Ve stavebnictví se naopak proces standardizace často ubírá opačným směrem - vznikají nové formáty popisu dat, nové standardy a nové koncepce validace dat, které ne vždy vedou ke skutečné jednotnosti a praktické použitelnosti. Výsledkem je, že namísto zjednodušení a automatizace zpracování se odvětví potýká s dalšími úrovněmi regulace a byrokracie (obr. 6.3-1), což ne vždy přispívá ke zvýšení efektivity.



Obr. 6.3-1 Požadavky na obsah dat a informací jsou redukovány na popis atributů a jejich hraničních hodnot, popsaných pomocí tabulek.

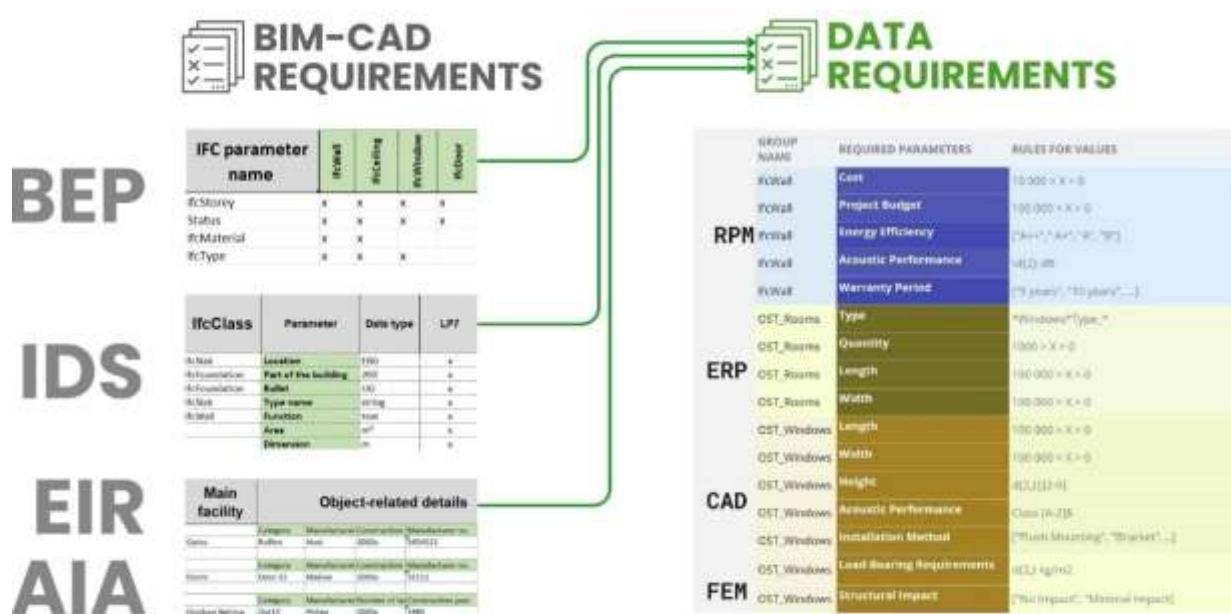
Nové koncepty související s daty CAD (BIM) namísto zjednodušení zpracování dat nejčastěji vyvolávají další složitosti a spory již ve fázi výkladu a základních definic.

Jedním z nejnovějších příkladů nových koncepcí je formát IDS (představený v roce 2020), který umožňuje popsat požadavky na složení atributů informačního modelu v otevřené koncepci BIM. Požadavky IDS popisují informace o attributech a jejich hraničních hodnotách ve formě strukturované tabulky (Excel nebo MySQL), která je následně převedena do značek polostrukturovaného formátu XML, přejmenovaného z XML na speciální zkratku IDS.

Na rozdíl od názoru prosazovaného dodavateli a podporovaného BIM a otevřeným BIM, že nakládání s daty ve stavebnictví je jedinečné kvůli používání specializovaných nástrojů, jako je CAD a BIM, se datové formáty a postupy správy dat v tomto odvětví neliší od postupů jiných odvětvích.

Počet požadavků na projekty a formáty CAD (BIM) lze zjednodušit použitím jediné tabulky požadavků s atributovými sloupcí, podrobně popsané v kapitole "Převod požadavků do strukturované podoby", aniž by bylo nutné převádět původně strukturované požadavky do netabulkových formátů (IDS je původně popsán prostřednictvím tabulky).

Zjednodušený přístup (obrázek 6.3-2), který zahrnuje sloupce pro identifikátory entit, vlastnosti a hraniční hodnoty, které byly podrobně rozebrány v předchozích kapitolách (obrázek 4.4-9, obrázek 4.4-16, obrázek 7.3-10), eliminuje potřebu převádět požadavky do formátu IDS-XML. Tato metoda poskytuje přímý, méně těžkopádný a transparentnější mechanismus kontroly kvality dat. Opírá se o široce používané nástroje, od regulárních výrazů (RegEx) po datové rámce, Pandas a standardní ETL - payplanes - přesně takové, jaké používají odborníci v jiných odvětvích ekonomiky pro práci s daty.



Obr. 6.3-2 Požadavky na data v jiných odvětvích jsou zjednodušeny na strukturovaný popis atributů a jejich hraničních hodnot.

Postupem času se ve stavebnictví vzhledem k uzavřené povaze dat objevuje stále více nových přístupů a technik pro kontrolu a správu těchto různorodých dat, ačkoli data ve stavebních projektech jsou v podstatě stejná jako v jiných oborech. Zatímco jiná odvětví si úspěšně vystačila se standardizovanými přístupy ke zpracování dat, ve stavebnictví se stále vyvíjejí nové a jedinečné datové formáty, požadavky a koncepce validace.

Metody a nástroje používané pro sběr, přípravu a analýzu údajů ve stavebnictví by se neměly zásadně lišit od metod a nástrojů používaných odborníky v jiných odvětvích hospodářství.

V tomto odvětví se vyvinul odlišný terminologický ekosystém, který vyžaduje kritickou reflexi a přehodnocení:

- Formát STEP je umístěn pod novým názvem IFC, doplněný o kategorizaci konstrukcí, aniž by byla zohledněna omezení samotného formátu STEP.
- Parametrický formát IFC se používá v procesech datové komunikace, přestože chybí jednotné geometrické jádro potřebné pro vizualizaci a výpočty.
- Přístup k databázím CAD -systémům je propagován pod pojmem "BIM", aniž by se hovořilo o specifikách těchto databází a přístupu k nim.
- Prodejci propagují interoperabilitu prostřednictvím formátů IFC a USD, často bez jejich uvedení do praxe, za použití nákladného reverzního inženýrství, se kterým se sami potýkají.
- Termíny LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD se používají univerzálně k popisu stejných parametrů entit, bez odkazu na nástroje modelování a ověřování, které se již dlouho používají v jiných odvětvích.

Stavebnictví ukazuje, že vše výše uvedené, i když to zní podivně, je ve stavebnictví možné - zejména pokud je hlavním cílem zpěněžit každou fázi zpracování dat prostřednictvím prodeje specializovaných služeb a softwaru. Z obchodního hlediska na tom není nic špatného. Otevřenou otázkou však zůstává, zda takové zkratky a přístupy spojené s CAD (BIM) skutečně přináší přidanou hodnotu a zjednoduší odborné procesy.

Ve stavebnictví takový systém funguje, protože samo odvětví v tomto bludišti systémů a zkratek dosahuje většiny svých spekulativních zisků. Společnosti, které mají zájem o transparentní procesy a otevřená data, jsou vzácné. Tato složitá situace bude pravděpodobně trvat donekonečna - dokud zákazníci, klienti, investoři, banky a soukromý kapitál nezačnou požadovat jasnější a informovanější přístupy ke správě informací.

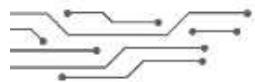
V oboru se nashromáždilo nadměrné množství zkratek, ale všechny různé míře popisují stejné procesy a požadavky na data. Jejich skutečná užitečnost pro zjednodušení pracovních postupů zůstává sporná.

Zatímco pojmy a marketingové zkratky přicházejí a odcházejí, samotné procesy validace datových požadavků zůstanou navždy nedílnou součástí obchodních procesů. Místo toho, aby vytvářel další a další specializované formáty a předpisy, by se stavebnictví mělo poohlédnout po nástrojích, které se již osvědčily v jiných oblastech, jako jsou finance, průmysl a IT.

Množství termínů, zkratek a formátů vytváří iluzi hluboce propracovaných digitálních konstrukčních procesů. Marketingové pojmy a složitá terminologie však často skrývají jednoduchou, ale nepříjemnou pravdu: data zůstávají těžko dostupná, špatně zdokumentovaná a pevně vázaná na konkrétní softwarová řešení.

Abychom se dostali z tohoto začarovaného kruhu zkratek a formátů pro formáty, je třeba se na

systémy CAD (BIM) dívat nikoli jako na magické nástroje pro správu informací, ale jako na to, čím skutečně jsou - specializované databáze. A právě tímto prizmatem lze pochopit, kde končí marketing a začíná skutečná práce s informacemi.



KAPITOLA 6.4.

PARAMETRIZACE NÁVRHU A POUŽITÍ LLM PRO PROVOZ V SYSTÉMU CAD

Iluze jedinečnosti dat CAD (BIM): cesta k analytice a otevřeným formátům

Moderní platformy CAD (BIM) významně změnily přístup k projektování a správě informací o stavbě. Zatímco dříve tyto nástroje sloužily především k vytváření výkresů a 3D modelů, dnes fungují jako plnohodnotná úložiště projektových dat. V rámci koncepce jediného zdroje pravdy se parametrický model stále více stává hlavním a často jediným zdrojem projektových informací, což zajíšťuje jejich integritu a relevanci po celou dobu životního cyklu projektu.

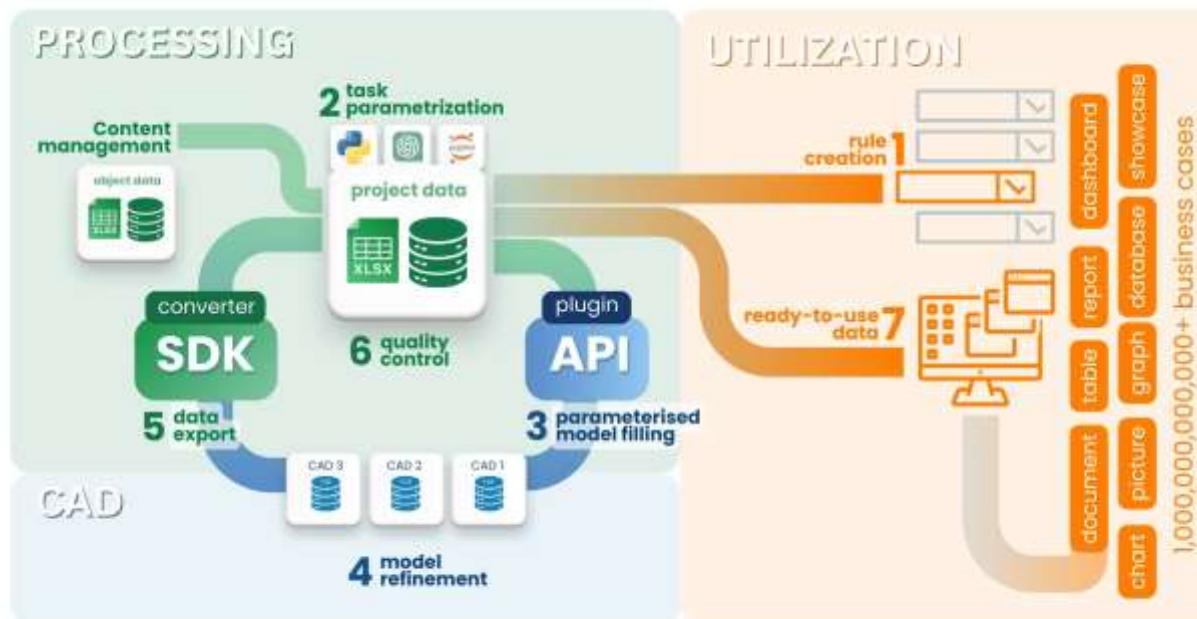
Klíčovým rozdílem mezi platformami CAD - (BIM-) a ostatními systémy pro správu stavebních dat je potřeba specializovaných nástrojů a rozhraní API pro přístup k informacím (jediný zdroj pravdy). Tyto databáze nejsou univerzální v tradičním smyslu: místo otevřené struktury a flexibilní integrace představují uzavřené prostředí, pevně svázané s konkrétní platformou a formátem.

Navzdory složitosti práce s daty CAD -data je důležitější otázka, která přesahuje technickou realizaci: co jsou vlastně databáze CAD (BIM)? Pro zodpovězení této otázky je nutné jít nad rámec obvyklých zkratek a pojmu vnučených tvůrce softwaru. Místo toho je vhodné zaměřit se na podstatu práce s projektovými informacemi: data a jejich zpracování.

Obchodní proces ve stavebnictví nezačíná prací v nástrojích CAD - nebo BIM -, ale tvorbou požadavků na projekt a datovým modelováním. Nejprve se definují parametry úlohy: seznam entit, jejich výchozí charakteristiky a hraniční hodnoty, které je třeba zohlednit při řešení konkrétní úlohy. Teprve poté se v systémech CAD (BIM) na základě zadaných parametrů vytvázejí modely a prvky

Proces, který předchází vytvoření informací v databázích CAD - (BIM-), je zcela shodný s procesem datového modelování, který byl podrobně rozebrán ve čtvrté části knihy a v kapitole "Datové modelování: konceptuální, logický a fyzický model" (obr. 4.3-1).

Stejně jako při datovém modelování vytváříme požadavky na data, která chceme později zpracovávat v databázi, vytvářejí manažeři databází CAD požadavky na návrh v podobě několika sloupců tabulky nebo seznamů dvojic klíč-hodnota (obr. 6.4-1, kroky 1-2). A teprve na základě těchto výchozích parametrů pomocí API automaticky nebo ručně projektant vytváří (nebo spíše upřesňuje) objekty v CAD- (BIM) databázích (kroky 3-4), načež je opět kontroluje na shodu s výchozími požadavky (kroky 5-6). Tento proces - definice → tvorba → validace → úprava (kroky 2-6) - se opakuje iterativně, dokud kvalita dat, stejně jako při datovém modelování, nedosáhne požadované úrovně pro cílový systém - dokumenty, tabulky nebo informační panely (krok 7).

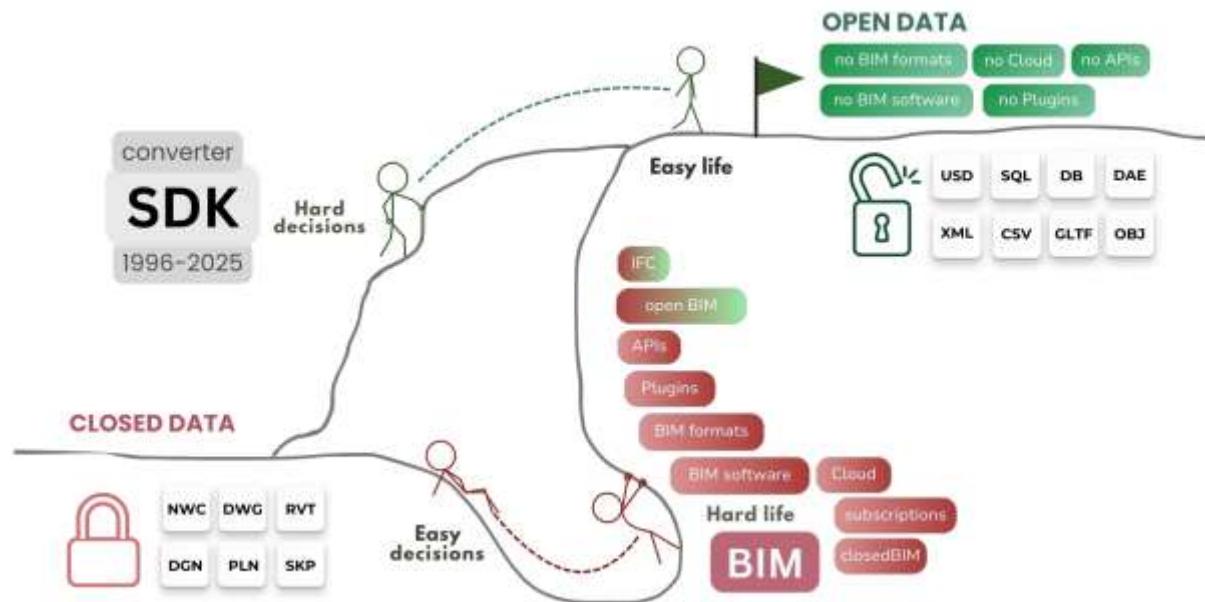


Obr. 6.4-1 Cyklus informační nasycenosti databází pro podnikové procesy při realizaci stavebních projektů.

Pokud budeme považovat CAD (BIM) za mechanismus přenosu parametrů ve formě souboru dvojic klíč-hodnota generovaných na základě požadavků definovaných mimo návrhové prostředí (obr. 6.4-1, kroky 1-2), těžiště diskuse se přesune od konkrétních softwarových řešení a jejich omezení k zásadnějším aspektům - datové struktuře, datovým modelům a datovým požadavkům. V podstatě hovoříme o parametrické saturaci databáze a klasickém procesu datového modelování (kroky 2-3 a 5-6). Jediným rozdílem je, že vzhledem k uzavřené povaze CAD-databází a zvláštnostem používaných formátů je tento proces doprovázen použitím specializovaných BIM-nástrojů. Vyhstává otázka: v čem spočívá jedinečnost BIM, když v jiných odvětvích podobné přístupy neexistují?

Za posledních 20 let se BIM stal více než jen jediným zdrojem dat. Svazek CAD -BIM je často prodáván jako parametrický nástroj s inherentně integrovanou databází [64], který je schopen automatizovat procesy navrhování, modelování a řízení životního cyklu stavebních projektů. Ve skutečnosti se však BIM stal spíše nástrojem pro udržení uživatelů na platformě dodavatele než vhodnou metodou správy dat a procesů.

Výsledkem je, že data CAD- (BIM-) jsou v rámci jejich platform izolována a informace o projektu se skrývají za proprietárními rozhraními API a geometrickými jádry. Uživatelé tak přišli o možnost nezávislého přístupu k databázím a o možnost získávat, analyzovat, automatizovat a přenášet data do jiných systémů, a obejít tak ekosystémy dodavatelů.



Obr. 6.4-2 Ve stavebnictví vyžadují moderní formáty sofistikovaná geometrická jádra, každoročně aktualizované API a speciální licence pro CAD -(BIM-) software.

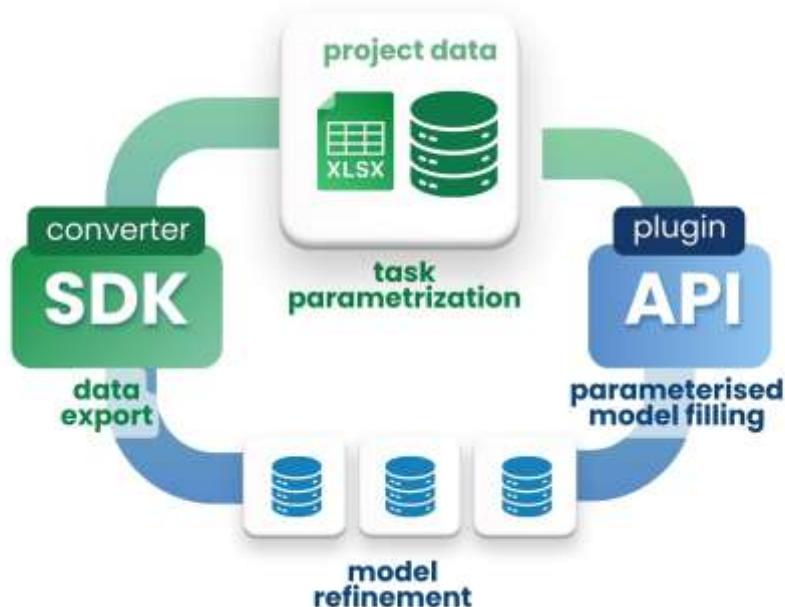
Společnosti, které pracují s moderními nástroji CAD, by měly používat stejný přístup k práci s daty, který v praxi používají všichni samotní dodavatelé CAD bez výjimky: transformaci dat pomocí SDK - nástroje reverzního inženýrství, proti kterému dodavatelé CAD bojují již od roku 1995 [75]. Máme-li plný přístup k databázi CAD a používáme-li nástroje reverzního inženýrství, můžeme získat [127] plochou množinu entit s atributy a exportovat je do libovolného vhodného otevřeného formátu (obr. 6.4-2), včetně geometrie i parametrů konstrukčních prvků. Tento přístup zásadně mění paradigmum práce s informacemi - od souborově orientované architektury k datově orientované:

- Datové formáty jako RVT, IFC, PLN, DB1, CP2, CPIXML, USD, SQLite, XLSX, PARQUET a další obsahují shodné informace o prvcích téhož projektu. To znamená, že znalost konkrétního formátu a jeho schématu by neměla být překážkou pro práci se samotnými daty.
- Data z libovolných formátů lze kombinovat do jediné otevřené strukturované a granulární struktury (obrázek 9.1-10), která obsahuje trojúhelníkovou geometrii MESH a vlastnosti všech objektových entit bez omezení geometrických jader.
- Analýza dat usiluje o univerzálnost: pomocí otevřených dat můžete pracovat s daty projektu bez ohledu na použitý formát.
- Minimalizace a závislost na rozhraní API a zásuvných modulech dodavatelů: práce s daty již není závislá na znalostech rozhraní API.

Když se požadavky na a CAD -data transformují do snadno analyzovatelných formátů strukturované reprezentace - vývojáři již nejsou závislí na specifických datových schématech a uzavřených ekosystémech.

Navrhování pomocí parametrů: budoucnost CAD a BIM

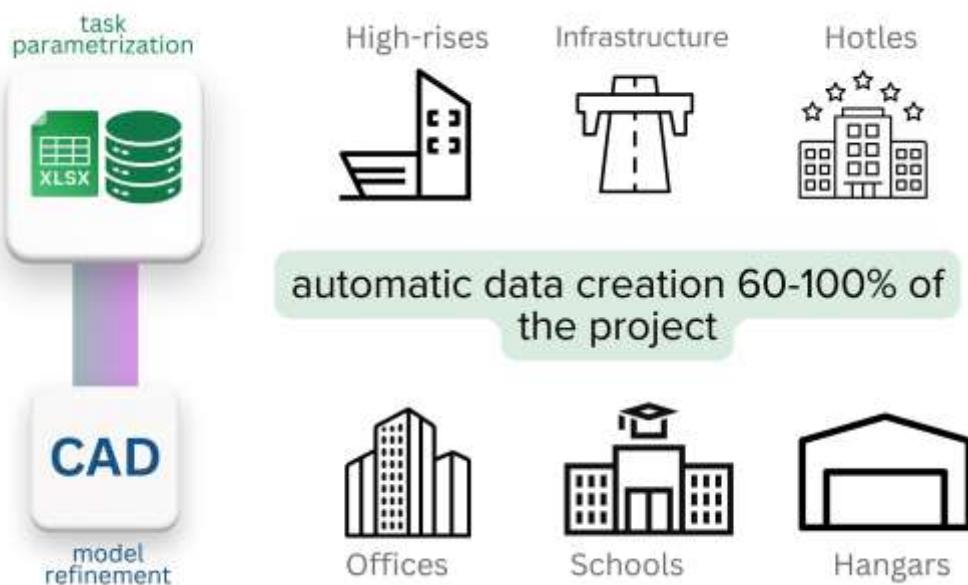
Žádný stavební projekt na světě ještě nezačal v programu CAD. Než se výkres nebo model v CADu zformuje, projde fází konceptualizace (obr. 6.4-1, fáze 1-2), kde se pozornost soustředí na parametry, které definují základní myšlenku a logiku budoucího objektu. Tato fáze odpovídá konceptuální úrovni v datovém modelování (obr. 4.3-6). Parametry mohou existovat pouze v mysli projektanta, ale v ideálním případě jsou usporádány ve formě strukturovaných seznamů, tabulek nebo uloženy v databázích (obr. 6.4-3), což umožňuje přehlednost, reprodukovatelnost a další automatizaci procesu návrhu.



Obr. 6.4-3 Proces návrhu je iterativní proces doplňování databáze CAD informacemi zvenčí s využitím požadavků v hodnotovém řetězci.

Před zahájením samotného modelování v CAD (logická a fyzická fáze datového modelování (obr. 4.3-7)) je důležité definovat hraniční parametry, které slouží jako základ projektu. Tyto atributy se stejně jako ostatní požadavky shromažďují na samém konci řetězce využití dat (např. systémů) a jejich prostřednictvím jsou již definována omezení, cíle a klíčové vlastnosti budoucích objektů v projektu.

Samotné modelování lze plně automatizovat z 60-100% pomocí nástrojů pro parametrické modelování (obr. 6.4-3), pokud jsou dobře definovány požadavky. Jakmile je projekt popsán ve formě parametrů, stává se jeho tvorba technicky proveditelnou, např. pomocí vizuálních programovacích jazyků, jako je Grasshopper Dynamo, zabudovaných v moderních CAD -prostředích nebo bezplatných řešení v Blenderu, UE, Omniverse.



Obr. 6.4-4 Většina typizovaných projektů je dnes již vytvářena plně automaticky díky parametrickým programátorským nástrojům.

Velké průmyslové a typizované projekty již dnes nevznikají rukama designérů, ale pomocí parametrických nástrojů a vizuálního programování. To umožňuje vytvořit model založený na datech, a nikoli na subjektivním rozhodnutí konkrétního projektanta nebo manažera.

Obsah předchází designu. Design bez obsahu není design, ale dekorace [128].

- Jeffrey Zeldman, webový designér a podnikatel

Proces nezačíná výkresem nebo 3D -modelováním, ale tvorbou požadavků. Právě požadavky určují, jaké prvky budou v projektu použity, jaká data je třeba přenést do jiných oddělení a systémů. Teprve existence strukturovaných požadavků umožňuje automatickou pravidelnou kontrolu modelů (například i každých 10 minut, aniž by to projektanta odvádělo od práce).

Možná, že v budoucnu se systém CAD- (BIM-stane pouze rozhraním pro naplnění databáze a nebude záležet na tom, v jakém nástroji CAD modelování provádí (fyzická úroveň).

Podobně ve strojírenství se často používá 3D modelování, které však není nezbytným nebo povinným prvkem projektu. Ve většině případů postačí klasická 2D dokumentace a na jejím základě se vytvoří potřebný informační model. Tento model je sestaven z komponent strukturovaných podle průmyslových standardů a obsahuje všechny potřebné informace pro pochopení konstrukce a organizace výroby. Z informačního modelu závodu se pak vytvoří informační model závodu, do kterého se doplní konkrétní výrobky a vývojové diagramy, které jsou již orientovány na potřeby technologů. Celý proces lze organizovat bez zbytečné složitosti, aniž by se systém zatěžoval 3D grafikou tam, kde

nepřináší skutečné výhody.

Je důležité si uvědomit, že samotný 3D model a CAD -systém by neměl hrát hlavní roli - je to pouze nástroj pro kvantitativní a geometrickou analýzu. Všechny ostatní parametry, kromě geometrie, které popisují entitu, by měly být uloženy a zpracovány pokud možno mimo prostředí CAD (BIM).

Navrhování podle parametrů není jen trendem, ale nevyhnutelnou budoucností stavebnictví. Místo ručního vytváření složitých 3D -modelů budou projektanti pracovat s daty, ověřovat je a automatizovat procesy, čímž se stavebnictví přiblíží světu programování. Postupem času budou projekční procesy postaveny na principech vývoje softwaru:

- Vytvoření požadavků → Vytvoření modelu → Nahrání na server → Ověření změn → Pull request
- Požadavek Pull automaticky provede kontrolu modelu na základě požadavků, které byly vytvořeny před procesem návrhu nebo v jeho průběhu.
- Po kontrole kvality dat a schválení jsou změny implementovány do projektu, společné databáze nebo automaticky přeneseny do jiných systémů.

Již nyní se ve strojírenství takové změny návrhu zahajují vytvořením oznámení o změně. Podobný systém čeká i stavebnictví: projektování bude iterativním procesem, kde každý krok bude podložen parametrickými požadavky. Takový systém umožní projektantům vytvářet automatické kontroly a automatické žádosti o stažení konkrétních požadavků.

Projektant budoucnosti je především operátor dat, nikoliv manuální modelář. Jeho úkolem je naplnit projekt parametrickými entitami, kde geometrie je pouze jedním z atributů.

Právě pochopení významu datového modelování, klasifikace a standardizace, které byly podrobně rozebrány v předchozích kapitolách knihy, bude hrát při transformaci důležitou roli. Návrhové předpisy budoucnosti budou formalizovány jako dvojice klíč-hodnota parametrů ve formě XLSX nebo XML - schémat.

Budoucnost stavebnictví je ve sběru dat, jejich analýze, ověřování a automatizaci procesů pomocí analytických nástrojů. BIM (nebo CAD) není konečným cílem, ale pouze etapou vývoje. Až si odborníci uvědomí, že mohou pracovat přímo s daty a obejít tradiční nástroje CAD, samotný pojem "BIM" postupně ustoupí pojmu využívání strukturovaných a granulovaných dat stavebního projektu.

Jedním z klíčových faktorů urychlujících tuto transformaci byl vznik velkých jazykových modelů (LLM) a nástrojů na nich založených. Tyto technologie mění způsob, jakým se pracuje s daty o návrhu, a umožňují přístup k informacím bez nutnosti hluboké znalosti rozhraní API nebo řešení dodavatele. Díky LLM se proces vytváření požadavků a interakce s daty CAD stává intuitivním a přístupným.

Vznik LLM v procesech zpracování dat v návrhovém CAD

Vedle vývoje nástrojů pro přístup k databázím CAD a otevřených a zjednodušených formátů CAD přinesl revoluci ve zpracování návrhových dat také vznik nástrojů LLM (Large Language Models). Zatímco v minulosti se k informacím přistupovalo převážně prostřednictvím složitých rozhraní a vyžadovalo to programátorské dovednosti a znalost API, nyní je možné s daty komunikovat pomocí přirozeného jazyka.

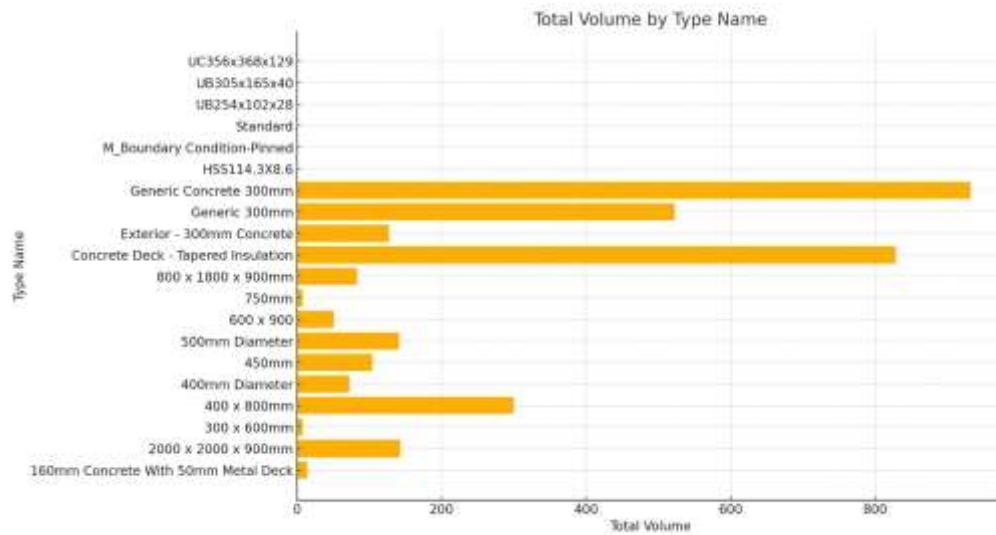
Inženýři, manažeři a projektanti bez technického vzdělání mohou získat potřebné informace z projektových dat formulováním dotazů v běžném jazyce. Za předpokladu, že jsou data strukturovaná a přístupná (obr. 4.1-13), stačí položit dotaz v LLM chat typu: "Zobrazte v tabulce se seskupením podle typu všechny stěny s objemem větším než 10 metrů krychlových" - a model tento dotaz automaticky převede do jazyka SQL nebo kódu v Pandas, čímž vygeneruje souhrnnou tabulkou, graf nebo dokonce hotový dokument.

Níže uvádíme několik reálných příkladů interakce LLM -modelů s návrhovými daty reprezentovanými v různých formátech CAD- (BIM-).

- 💡 Příklad dotazu v LLM chatu na projekt CAD ve formátu RVT po převodu (obr. 4.1-13) do tabulkového datového rámce (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN nebo jiného):

Seskupit data v datovém rámci získaná ze souboru RVT podle "názvu typu" při sčítání parametru "Objem" a zobrazit počet prvků ve skupině. A to vše prosím zobrazte jako horizontální histogram bez nulových hodnot.

- 💡 Odezva LLM jako horizontální sloupcový graf (formát PNG):



Obrázek 6.4-5 Místo 17 kliknutí myší nebo 40 řádků kódu pomocí zásuvných modulů v LLM okamžitě načteme tabulku QTO pomocí textového dotazu.

- 💡 Pro vygenerování tabulky QTO typů stěn s celkovou plochou a množstvím z kategorie "Stěny" formulujme textový dotaz pro LLM -chat:

Vezměte z datového rámce projektu pouze ty položky, které mají v parametru "Kategorie" hodnotu "OST_Stěny", seskupte je podle "Názvu typu", sečtěte hodnotu sloupce "Plocha", přičtěte množství a zobrazte je v tabulce odstraněním nulových hodnot.

- 💡 Odpověď LLM ve formě hotové tabulky QTO:

Type Name	Total Area	Count
CL_W1	393.12 sq m	10
Cavity wall_sliders	9.37 sq m	1
Foundation - 300mm Concrete	30.90 sq m	1
Interior - 165 Partition (1-hr)	17.25 sq m	3
Interior - Partition	186.54 sq m	14
Retaining - 300mm Concrete	195.79 sq m	10
SH_Curtain wall	159.42 sq m	9
SIP 202mm Wall - conc clad	114.76 sq m	4
Wall - Timber Clad	162.91 sq m	8

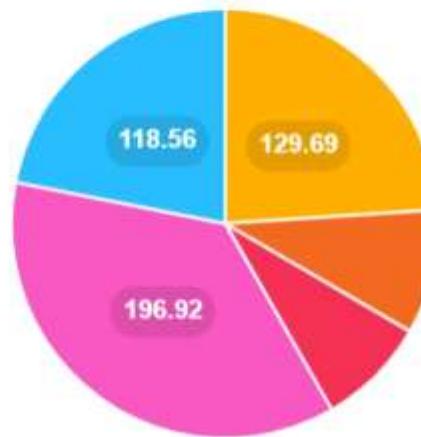
Obr. 6.4-6 Vytvoření tabulky QTO v přirozeném jazyce poskytuje stejně kvalitní výsledek jako při použití nástrojů CAD - (BIM-).

- 💡 Dotazujme se na projekt ve formátu IFC po převodu na tabulkový datový rámec a zadejme podobný textový dotaz do libovolného chatu LLM:

Vezměte z projektu pouze ty položky, které mají v parametru "Rodíč" hodnoty Úroveň 1 a Úroveň 2, a vezměte položky, které mají v parametru "Kategorie" hodnoty IfcSlab, pak tyto položky seskupte podle parametru "Typ objektu", sečtěte hodnoty v parametru "PSet_RVT _Dimensions Area" a zobrazte je jako koláčový graf.

💡 Odpověď LLM jako hotový koláčový graf skupin prvků z dat IFC :

For Floor:127mm Slab on Grade, Floor:150mm Exterior Slab on Grade,
 Floor:Finish Floor - Ceramic Tile, Floor:Finish Floor - Wood, and
 Floor:Residential - Wood Joist with Subflooring



Obrázek 6.4-7 Výsledkem dotazu na data IFC ve strukturovaném formátu může být jakýkoli typ grafu, který je vhodný pro pochopení dat.

Za každým z výsledných hotových řešení (obr. 6.4-5 - obr. 6.4-7) se skrývá několik desítek řádků kódu Pythonu využívajícího knihovnu Pandas. Výsledný kód lze zkopirovat z chatu LLM a použít v libovolném lokálním nebo online IDE a získat tak identické výsledky i mimo chat LLM.

Ve stejném chatu LLM můžeme pracovat nejen s projekty získanými z 3D CAD (BIM) formátů, ale také s plochými výkresy ve formátu DWG, na které se můžeme dotazovat v chatu LLM a zobrazit například data o skupinách prvků ve formě čar nebo 3D geometrie po převodu do strukturované podoby.

Automatizovaná analýza souborů DWG pomocí LLM a Pandas

Zpracování dat ze souborů DWG - vzhledem k nestrukturované povaze informací - bylo vždy složitým úkolem, který vyžadoval specializovaný software a často ruční analýzu. S rozvojem umělé inteligence a nástrojů LLM, však bylo možné mnoho kroků tohoto, dnes většinou manuálního procesu, automatizovat. Uvažujme skutečnou Pipeline požadavků na LLM (v tomto příkladu ChatGPT) pro práci s výkresy DWG, které umožňují pracovat s projektem:

- Filtrování dat DWG podle vrstvy, ID a souřadnic
- Vizualizace geometrie prvků
- Automatické anotování výkresů na základě parametrů
- Rozšíření polylinií stěn do vodorovné roviny
- Vytváření interaktivních 3D -vizualizace roviných dat

■ Strukturovat a analyzovat konstrukční data bez složitých nástrojů CAD.

V našem případě začíná proces sestavování Pipeline sekvenčním generováním kódu prostřednictvím LLM. Nejprve se vygeneruje dotaz popisující úlohu. ChatGPT vygeneruje kód v jazyce Python -kód, který se provede a analyzuje a zobrazí výsledek uvnitř chatu. Pokud výsledek neodpovídá očekávání, dotaz se opraví a proces se opakuje

Pipeline je posloupnost automatizovaných kroků prováděných za účelem zpracování a analýzy dat. V takovém procesu každý krok přijímá data jako vstup, provádí transformace a předává výsledek dalšímu kroku.

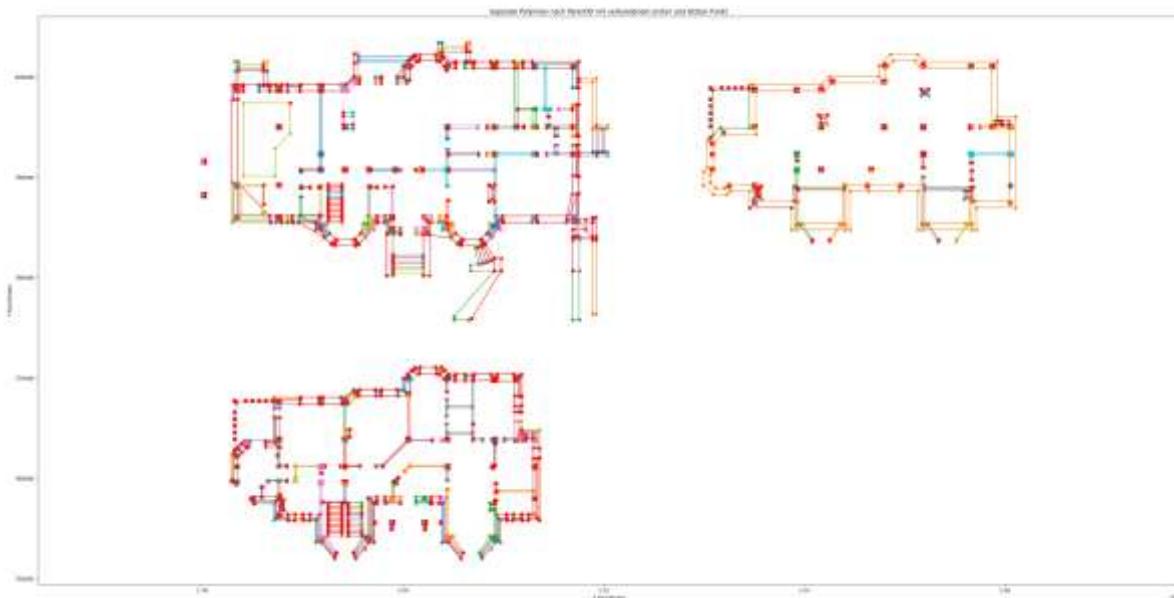
Po získání požadovaného výsledku se kód zkopíruje z LLM a vloží se do kódu ve formě bloků v některém z vhodných IDE, v našem případě na platformě Kaggle.com. Výsledné fragmenty kódu se spojí do jediné pipeline, která celý proces automatizuje - od načtení dat až po jejich finální analýzu. Tento přístup umožňuje rychlý vývoj a škálování analytických procesů bez hlubokých znalostí programování. Úplný kód všech níže uvedených fragmentů spolu s ukázkovými dotazy najeznete na platformě Kaggle.com po vyhledání "DWG Analyse with ChatGPT | DataDrivenConstruction". [129].

Začneme proces práce s DWG daty, po převodu do strukturované podoby (obr. 4.1-13), klasickým krokem - seskupením a filtrováním všech výkresových dat, potřebných pro naši úlohu stěnových prvků, konkrétně polylinií (parametr "ParentID" umožňuje seskupit linie do skupin), které mají v parametru (sloupci datového rámce) "Layer" řetězcovou hodnotu obsahující následující kombinaci písmen (RegEx) - "wall".

- Chcete-li získat kód pro podobnou úlohu a výsledek v podobě obrázku, měli byste v LLM napsat následující dotaz:

Nejprve zkontrolujte, zda datový rámec získaný z DWG obsahuje definované sloupce: 'Layer', 'ID', 'ParentID' a 'Point'. Poté vyfiltrujte ID ze sloupce "Layer", která obsahují řetězec "wall". Najděte položky ve sloupci "ParentID", které odpovídají těmto identifikátorům. Definujte funkci, která vyčistí a rozdělí data ve sloupci "Point". To zahrnuje odstranění závorek a rozdelení hodnot na souřadnice "x", "y" a "z". Vykreslete data pomocí matplotlib. Pro každé jedinečné "ParentID" nakreslete samostatnou polylinii spojující souřadnice "Point". Ujistěte se, že první a poslední bod jsou pokud možno propojeny. Nastavte příslušné popisky a nadpisy a zajistěte, aby osy x a y měly stejně měřítko.

- Odpověď LLM vám poskytne hotový obrázek, za kterým se skrývá kód Pythonu, který jej vygeneroval:



Obr. 6.4-8 Kód LLM extrahal ze souboru DWG všechny čáry vrstvy "zed", vymazal jejich souřadnice a zkonstruoval polylinie pomocí jedné z knihoven Pythonu.

- ❷ Nyní přidejme k liniím parametr plochy, který má každá polylinie ve svých vlastnostech (v jednom ze sloupců datového rámce):

Nyní získejte z každé polylinie pouze jedno "ParentID" - najděte toto ID ve sloupci "ID", vezměte hodnotu "Area", vydělte ji 1 000 000 a přidejte tuto hodnotu do grafu.

- V odpovědi LLM se zobrazí nový graf, kde každá polylinie bude mít popisek se svou plochou:

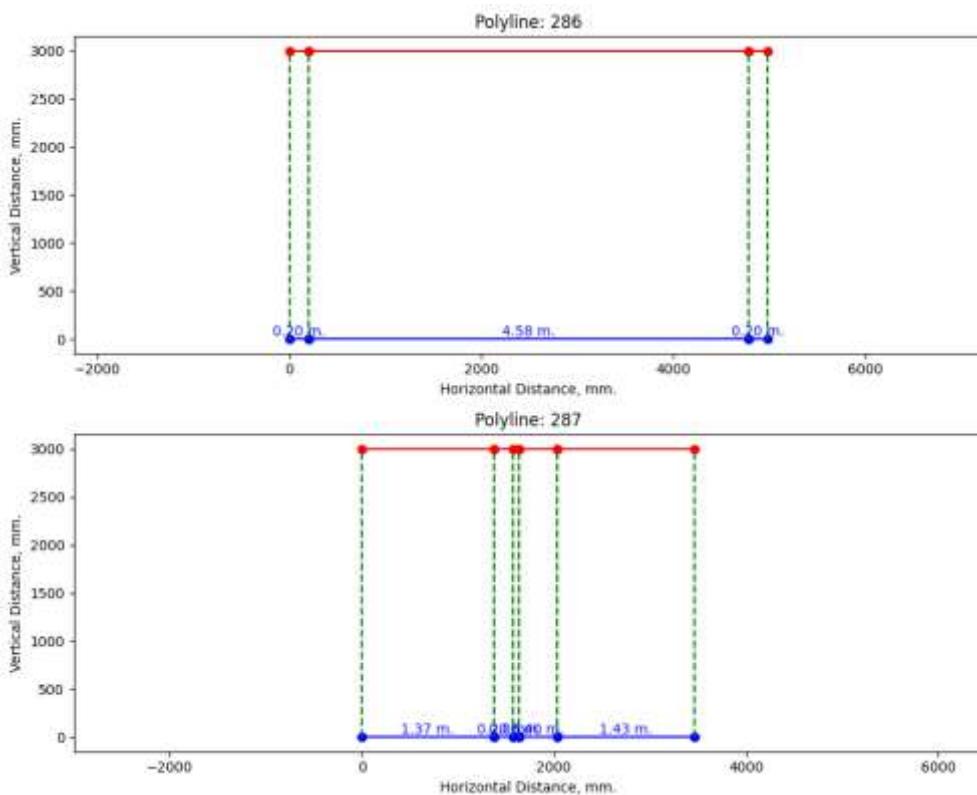


Obr. 6.4-9 LLM přidal kód, který přebírá hodnoty plochy pro každou polylinii a přidává je do obrázku s vizualizací linií.

- Poté převedeme každou polylinii na vodorovnou čáru, přidáme rovnoběžnou čáru ve výšce 3000 mm a spojíme je do jedné roviny, abychom tímto způsobem zobrazili rozložení ploch stěnových prvků:

Je třeba vzít všechny prvky ze sloupce "Layer" s hodnotou "wall". Vezměte tato ID jako seznam ze sloupce "ID" a najděte tato ID z celého datového rámce ve sloupci "ParentID". Všechny prvky jsou linie, které jsou spojeny do jedné polylinie. Každá linie má jinou geometrii x, y prvního bodu ve sloupci "Point". Musíte postupně vzít každou polylinku a z bodu 0,0 vodorovně vynést délku každé úsečky z polylinky. délku každé úsečky polylinky do jedné čáry. Pak nakreslete úplně stejně čáry, jen o 3000 vyšší, a spojte všechny body do jedné roviny.

- ❷ Odezva LLM vypisuje kód, který umožňuje vykreslit výkresy stěn v rovině:

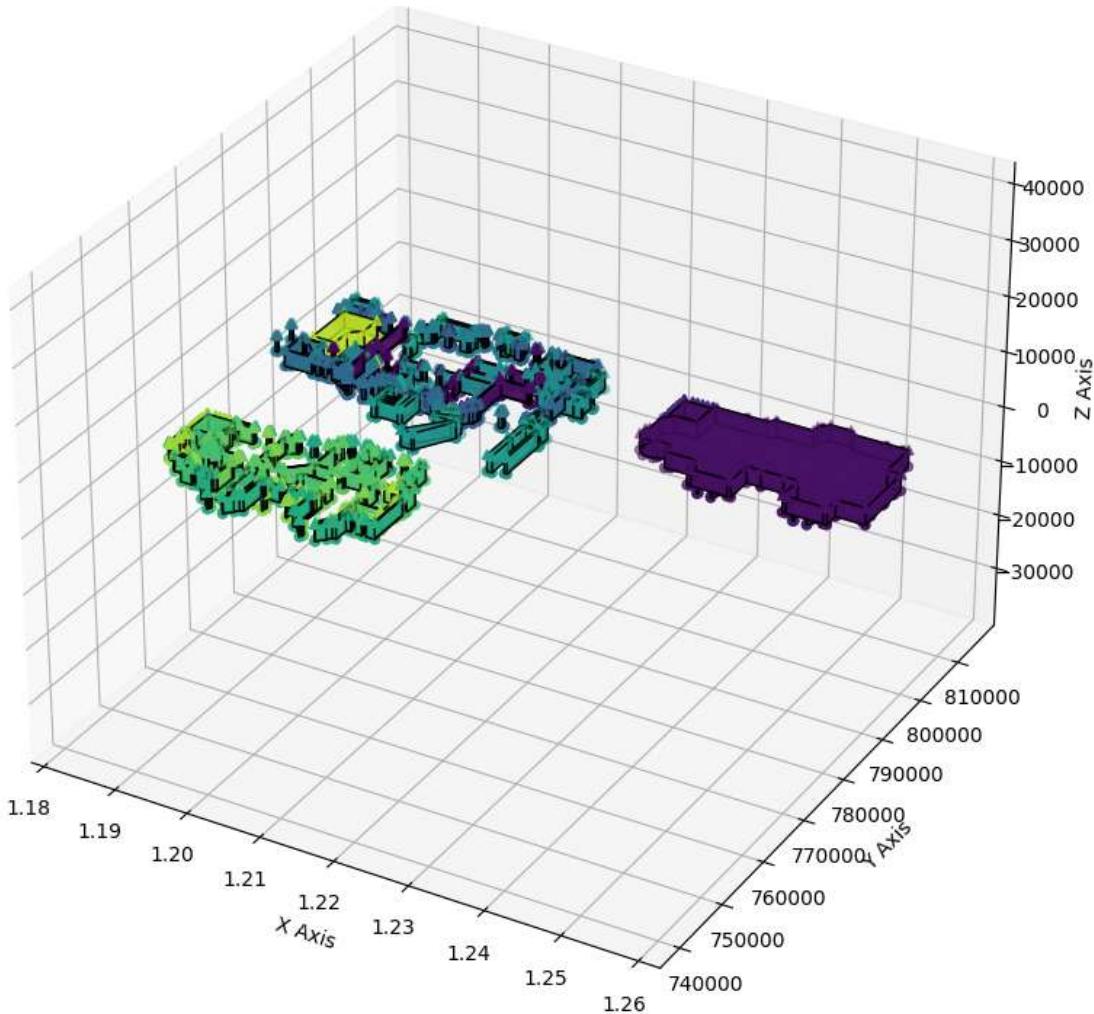


Obr. 6.4-10 Každou polylinii pomocí výzev změníme na layout, který vizualizuje roviny stěn přímo v chatu LLM.

- ❸ Nyní přejděme od 2D projekce k 3D - modelování stěn z plochých čar spojením horní a dolní vrstvy polylinií:

Vizualizujte prvky stěn ve 3D, spojte polylinie ve výškách $z = 0$ a $z = 3000$ mm.
Vytvoření uzavřené geometrie představující stěny budovy. Použijte 3D grafický nástroj Matplotlib.

- LLM vygeneruje interaktivní 3D -graf, ve kterém bude každá polylinie reprezentována jako sada rovin. Uživatel se bude moci volně pohybovat mezi prvky pomocí počítačové myši a zkoumat model ve 3D režimu kopírováním kódu z chatu do IDE:



Obr. 6.4-11 LLM pomohl vytvořit kód [129] pro vizualizaci plochých kreslicích čar do 3D zobrazení , které lze prozkoumat v 3D prohlížeči v prostředí IDE.

Pro vytvoření logické a reprodukovatelné potrubní linky - od počáteční konverze a načtení souboru DWG - až po konečný výsledek - se doporučuje po každém kroku zkopírovat vygenerovaný blok kódu LLM - do IDE. Tímto způsobem nejen zkontrolujete výsledek v chatu, ale také jej ihned spusťte ve svém vývojovém prostředí. To vám umožní postupně sestavovat proces, ladit jej a upravovat podle potřeby.

Kompletní kód Pipeline všech fragmentů (obrázky 6.4-8 až obrázky 6.4-11) spolu s ukázkovými dotazy najeznete na platformě Kaggle.com vyhledáním "DWG Analyse with ChatGPT | DataDrivenConstruction". [129]. Na serveru Kaggle si můžete nejen prohlédnout kód a použité výzvy,

ale také si celý Pipeline s původními datovými dataframy DWG zdarma zkopiřovat a otestovat v cloudu, aniž byste museli instalovat jakýkoli další software nebo samotné IDE.

Přístup uvedený v této kapitole umožnuje plně automatizovat kontrolu, zpracování a generování dokumentů na základě DWG -projektů. Vyvinutý Pipeline je vhodný jak pro zpracování jednotlivých výkresů, tak pro dávkové zpracování desítek, stovek a tisíců DWG-souborů s automatickým generováním potřebných sestav a vizualizací pro každý projekt.

Proces lze uspořádat sekvenčně a přehledně: nejprve se data ze souboru CAD automaticky převedou do formátu XLSX, poté se načtou do datového rámce, následuje seskupování, kontrola a generování výsledků - to vše se realizuje v jediném zápisníku Jupyter nebo ve skriptovacím jazyce Python v libovolném populárním prostředí IDE. V případě potřeby lze proces snadno rozšířit integrací se systémy pro správu projektové dokumentace: soubory CAD lze automaticky načítat podle zadaných kritérií, výsledky lze vracet zpět do úložného systému a uživatelé mohou být upozorněni, až budou výsledky připraveny - e-mailem nebo prostřednictvím messengerů.

Používání chatů a agentů LLM pro práci s daty návrhu snižuje závislost na specializovaných programech CAD a umožňuje provádět analýzu a vizualizaci architektonických návrhů bez nutnosti manuální interakce s rozhraním - bez klikání myši a pamatování si složité navigace v menu.

S každým dalším dnem bude ve stavebnictví stále více slyšet o LLM, granulárních strukturovaných datech, DataFrame a sloupcových databázích. Jednotné dvourozměrné DataFramy vytvořené z různých databází a CAD formátů, budou ideálním palivem pro moderní analytické nástroje, se kterými aktivně pracují specialisté v jiných odvětvích.

Samotný proces automatizace se výrazně zjednoduší - místo studování API uzavřených niche produktů a psaní složitých skriptů pro analýzu nebo transformaci parametrů nyní bude stačit formulovat úlohu v podobě sady jednotlivých textových příkazů, které se složí do požadovaného Pipeline nebo Workflow-procesu pro požadovaný programovací jazyk, který běží zdarma na téměř libovolném zařízení. Už žádné čekání na nové produkty, formáty, plug-iny nebo aktualizace od výrobců nástrojů CAD- (BIM-). Inženýři a stavitele budou moci samostatně pracovat s daty pomocí jednoduchých, bezplatných a snadno pochopitelných nástrojů, kterým budou asistovat chaty a agenti LLM.

Další kroky: přechod od uzavřených formátů k otevřeným datům

Při práci s nárhovými daty budoucnosti je nepravděpodobné, že by někdo skutečně potřeboval rozumět geometrickým jádrům proprietárních nástrojů nebo se učit stovek nekompatibilních formátů obsahujících stejně informace. Bez pochopení toho, proč je přechod k otevřeným strukturovaným datům důležitý, je však obtížné argumentovat ve prospěch používání nových svobodných nástrojů, otevřených dat a přístupů, které pravděpodobně nebudou prosazovány dodavateli softwaru.

V této kapitole jsme se zabývali hlavními vlastnostmi dat CAD (BIM), jejich omezeními a příležitostmi a tím, že navzdory marketingovým slibům dodavatelů se inženýři a projektanti každý den potýkají s

obtížemi při získávání, přenosu a analýze informací o návrhu. Pochopení architektury těchto systémů a seznámení se s alternativními přístupy - založenými na otevřených formátech a automatizaci prostřednictvím LLM - může výrazně usnadnit život i jednotlivým odborníkům, natož firmám. Pro shrnutí této části je vhodné zdůraznit hlavní praktické kroky, které vám pomohou aplikovat přístupy probírané v této části na vaše každodenní úkoly:

- Rozšiřte svou sadu nástrojů pro práci s projektovými daty
 - Prozkoumejte dostupné zásuvné moduly a nástroje pro získávání dat z používaných systémů CAD - (BIM-).
 - Prozkoumejte dostupné SDK a API, které umožňují automatizovat extrakci dat z uzavřených formátů bez nutnosti ručního otevírání specializovaného softwaru.
 - Zvládnutí základních dovedností práce s otevřenými neparametrickými geometrickými formáty (OBJ, glTF, USD, DAE) a odpovídajícími knihovnami s otevřeným zdrojovým kódem.
 - Zkuste vymyslet systém pro ukládání metadat projektu odděleně od geometrie mimo řešení CAD (BIM), abyste zjednodušili analýzu a integraci s dalšími systémy.
 - Použití LLM k automatizaci převodu dat mezi formáty
- Vytvoření vlastních procesů pro zpracování informací o projektu
 - začít popisovat úlohy a požadavky na modelování pomocí parametrů a jejich hodnot v jednoduchých a strukturovaných formátech.
 - Vytvoření osobní knihovny skriptů nebo bloků kódu pro často prováděné operace.
- Podporujte používání otevřených standardů ve své práci.
 - Vyzvat kolegy a partnery ke sdílení dat v otevřených formátech, které nejsou omezeny ekosystémem dodavatelů softwaru.
 - Ukázat výhody používání strukturovaných dat na konkrétních příkladech.
 - iniciovat diskusi o problémech s uzavřenými formáty a o možných řešeních.

I když nemůžete změnit politiku vaší společnosti týkající se platem CAD - (BIM-), osobní pochopení zásad práce s projektovými daty v otevřených formátech vám umožní výrazně zvýšit efektivitu vaší práce. Vytvořením vlastních nástrojů a metod pro extrakci a transformaci dat z různých formátů nejen optimalizujete své pracovní postupy, ale také získáte flexibilitu, která vám umožní obejít omezení standardních softwarových řešení.



VII ČÁST

ROZHODOVÁNÍ ZALOŽENÉ NA DATECH, ANALYTICE, AUTOMATIZACI A STROJOVÉM UČENÍ.

Sedmá část se zaměřuje na analýzu dat a automatizaci procesů ve stavebnictví. Pojednává o tom, jak se data stávají základem pro rozhodování, a vysvětluje zásady vizualizace informací pro efektivní analýzu. Podrobně jsou popsány klíčové ukazatele výkonnosti (KPI), metody vyhodnocování návratnosti investic (ROI) a vytváření informačních panelů pro sledování projektů. Zvláštní pozornost je věnována procesům ETL (Extract, Transform, Load) a jejich automatizaci pomocí potrubí (Pipeline), která mění různorodá data na strukturované informace pro analýzu. Diskutovány jsou nástroje pro orchestraci workflow, jako jsou Apache Airflow, Apache NiFi a n8n, které umožňují vytvářet automatizované datové pipelines bez hlubších znalostí programování. Významnou roli hrají velké jazykové modely (LLM) a jejich využití pro zjednodušení analýzy dat a automatizaci rutinních úloh

KAPITOLA 7.1.

ANALÝZA DAT A ROZHODOVÁNÍ ZALOŽENÉ NA DATECH.

Po krocích shromažďování, strukturování, čištění a ověřování informací vznikl ucelený a analyzovatelný soubor dat. Předchozí části knihy se zabývaly systematizací a strukturováním různorodých zdrojů - od dokumentů PDF a textových záznamů z jednání až po modely CAD a geometrická data. Podrobně je popsán proces kontroly a sladěování informací s požadavky různých systémů a klasifikátorů, odstraňování duplicit a nesrovnalostí.

Všechny výpočty prováděné nad těmito daty (třetí, čtvrtá část knihy) - od jednoduchých transformací až po výpočty časových, nákladových a ESG ukazatelů (pátá část) - jsou agregovanými analytickými úlohami. Tvoří základ pro pochopení aktuálního stavu projektu, posouzení jeho parametrů a následné rozhodování. V důsledku toho se data v důsledku výpočtů mění ze souboru nesourodých záznamů ve zvládnutelný zdroj schopný odpovědět na klíčové obchodní otázky.

V předchozích kapitolách byly podrobně popsány procesy sběru dat a kontroly kvality pro použití v typických obchodních případech a procesech specifických pro stavebnictví. Analytika v tomto kontextu je v mnoha ohledech podobná aplikacím v jiných průmyslových odvětvích, ale má řadu specifických rysů.

V následujících kapitolách bude podrobně popsán proces analýzy dat, včetně automatizačních kroků od prvotního získání informací a jejich transformace až po následný přenos do cílových systémů a dokumentů. Nejprve bude představena teoretická část, která se zaměří na vybrané aspekty rozhodování založeného na datech. Poté bude v následujících kapitolách představena praktická část týkající se automatizace a konstrukce ETL -Pipeline.

Data jako zdroj při rozhodování

Rozhodování založené na datech je často opakující se proces a začíná systematickým shromažďováním informací z různých informačních zdrojů. Stejně jako v přírodním koloběhu jednotlivé datové prvky i celé informační systémy postupně padají do půdy - hromadí se ve firemních informačních úložištích (obrázek 1.3-2). Postupem času se tato data, podobně jako spadané listí a větve, přeměňují v cenný materiál. Mycelium datových inženýrů a analytiků organzuje a připravuje informace pro budoucí použití a mění opadaná data a systémy v cenný kompost, aby z nich vyrostly nové výhonky a nové systémy (obr. 1.2-5).

Trendy v širokém využívání analytiky v různých odvětvích znamenají začátek nové éry, kdy se práce s daty stává základem profesní činnosti (obr. 7.1-1). Pro odborníky ve stavebnictví je důležité, aby se těmito změnami přizpůsobili a byli připraveni přejít do nové éry - éry dat a analytiky

Ruční přesouvání dat mezi tabulkami a ruční provádění výpočtů se postupně stává minulostí a ustupuje automatizaci, analýze toku dat, analytice a strojovému učení. Tyto nástroje se stávají klíčovými prvky moderních systémů pro podporu rozhodování.

V knize společnosti McKinsey "Restart. průvodce překonáváním konkurence ve věku digitálních technologií a umělé inteligence" [130] je citován průzkum provedený v roce 2022 mezi 1 330 vrcholovými manažery z různých regionů, odvětví a funkčních oblastí [130] cituje průzkum provedený v roce 2022 mezi 1 330 vrcholovými manažery z různých regionů, odvětví a funkčních oblastí. Podle jeho výsledků 70% vedoucích pracovníků využívá pokročilou analytiku k vytváření vlastních nápadů a 50% zavádí umělou inteligenci ke zlepšení a automatizaci rozhodovacích procesů.



Obr. 7.1-1 Analýza dat a analytika je hlavním nástrojem pro zvýšení rychlosti rozhodování ve firmě.

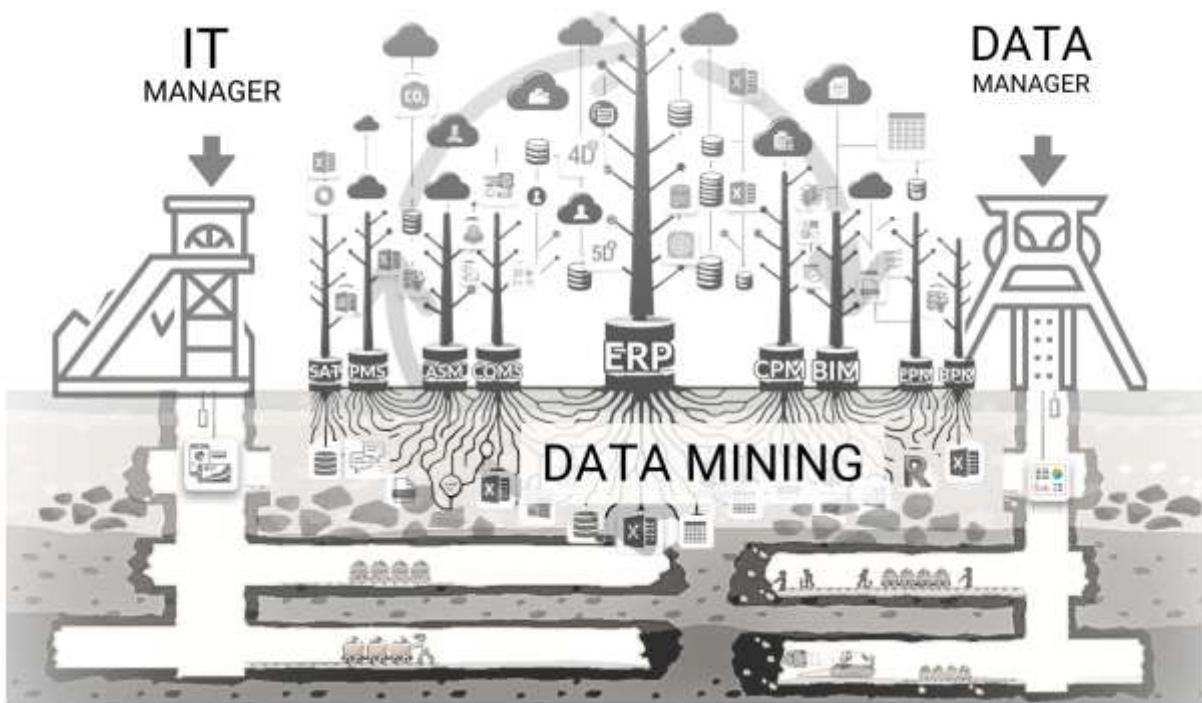
Analýza dat, podobně jako šíření mycelia, proniká do humusu minulých rozhodnutí, pomáhá propojovat jednotlivé systémy a vede manažery k cenným poznatkům. Tyto znalosti, podobně jako živiny z rozložitých stromů datových systémů, živí nová rozhodnutí ve firmě a vedou k efektivním změnám a růstu kvality informací, podobně jako nové výhonky a klíčky vyrůstající z bohaté a zdravé půdy (obrázek 1.2-5).

Čísla mají důležitý příběh. Spoléhají na to, že jim dáte jasný a přesvědčivý hlas [131].

- Stephen Few, odborník na vizualizaci dat

Ve středních a malých podnicích je dnes práce na získávání a přípravě informací pro další analýzu nesmírně pracný proces (obrázek 7.1-2), srovnatelný s těžbou uhlí v osmnáctém století. Donedávna byla práce na vytěžování a přípravě dat vyhrazena spíše dobrodruhům pracujícím ve vysoce specializované nice s malým a omezeným souborem nástrojů pro práci s různými typy dat z nestrukturovaných, volně strukturovaných, smíšených a uzavřených zdrojů.

Osoby s rozhodovacími pravomocemi a manažeři často nemají zkušenosti s heterogenními daty a systémy, ale musí na jejich základě přijímat rozhodnutí. V důsledku toho se rozhodování založené na datech v moderním stavebnictví v posledních desetiletích méně podobá automatizovanému procesu a spíše několikadenní ruční práci horníka v prvních uhelných dolech.



Obr. 7.1-2 V procesu dolování dat procházejí experti složitou cestou přípravy dat - od čištění až po strukturování pro následnou analýzu.

Ačkoli jsou moderní metody získávání dat ve stavebnictví jistě pokročilejší než primitivní techniky horníků z 12. století, stále se jedná o složitý a vysoce rizikový úkol, který vyžaduje značné zdroje a odborné znalosti, jež si mohou dovolit pouze velké společnosti. Procesy získávání a analýzy dat z nahromaděného dědictví minulých projektů donedávna prováděly převážně velké, technologicky vyspělé společnosti, které data soustavně shromažďovaly a uchovávaly po desetiletí

Dříve hrály vedoucí roli v oblasti analytiky technologicky vyspělé společnosti, které shromažďovaly data po desetiletí. Dnes se situace mění: přístup k datům a nástrojům pro jejich zpracování se demokratizuje - dříve složitá řešení jsou nyní dostupná všem zdarma.

Použití analytiky umožnuje společnostem přijímat přesnější a informovanější rozhodnutí v reálném čase. Následující případová studie ilustruje, jak mohou historická data pomoci při přijímání finančně správných rozhodnutí:

- ❑ **Projektový manažer** - "Nyní je průměrná cena betonu ve městě 82€ /m³, v odhadu máme 95 €/m³."
- ❑ **Odhadce** - "U předchozích projektů bylo překročení asi 15%, takže jsem se stáhl."
- ❑ **Manažer dat nebo inženýr kontroly na straně zákazníka** - "Podívejme se na analytiku posledních tří výběrových řízení."

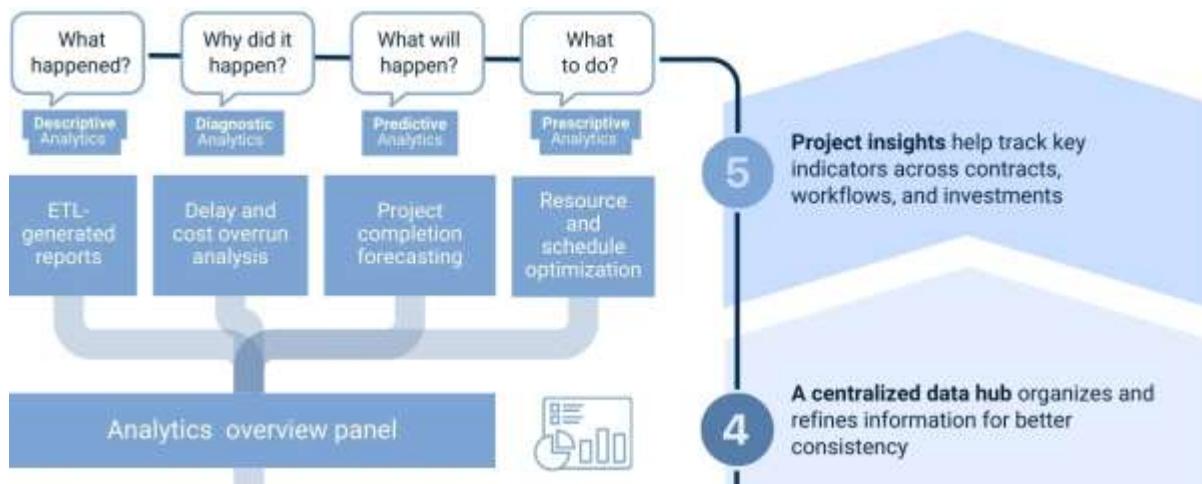
Po analýze DataFrame z minulých projektů získáme:

- **Průměrná skutečná nákupní cena:** 84,80 /m³

■ Průměrný poměr překročení výdajů: +4,7%

■ Doporučená sazba v odhadu: ~ 85 /m³€

Takové rozhodnutí již nebude založeno na subjektivních pocitech, ale na konkrétních historických statistikách, což pomáhá snižovat rizika a zvyšovat platnost nabídky do výběrového řízení. Analýza dat z minulých projektů se stává jakýmsi "organickým hnojivem", z něhož klíčí nová, přesnější řešení.



Obrázek 7.1-3 Analýza dat odpovídá na tři klíčové otázky: co se stalo, proč se to stalo a co by se mělo dělat dál.

Lidé s rozhodovacími pravomocemi a manažeři se často potýkají s nutností pracovat s různorodými daty a systémy bez dostatečných technických znalostí. V takových situacích je vizualizace, jeden z prvních a nejdůležitějších kroků v analytickém procesu, klíčovou pomůckou pro pochopení dat. Umožňuje prezentovat informace ve vizuální a srozumitelné podobě.

Vizualizace dat: klíč k pochopení a rozhodování

V dnešním stavebnictví, kde se projektová data vyznačují složitostí a víceúrovňovou strukturou, hraje vizualizace klíčovou roli. Vizualizace dat umožňuje projektovým manažerům a inženýrům vizualizovat složité vzorce a trendy skryté ve velkých, různorodých objemech dat.

Vizualizace dat usnadňuje pochopení stavu projektu: přidělování zdrojů, trendy nákladů nebo spotřeby materiálu. Grafy a schémata zpřístupňují složité a suché informace a umožňují rychle identifikovat klíčové oblasti, kterým je třeba věnovat pozornost, a odhalit potenciální problémy.

Vizualizace údajů nejen usnadňuje interpretaci informací, ale je i zásadním krokem v analytickém procesu a informovaném rozhodování managementu, neboť pomáhá odpovědět na otázky "co se stalo?" a "jak se to stalo?" (obrázek 2.2-5). (Obrázek 2.2-5).

Grafika je vizuální nástroj pro řešení logických problémů [132].

- Jacques Bertin, "Grafika a grafické zpracování informací.

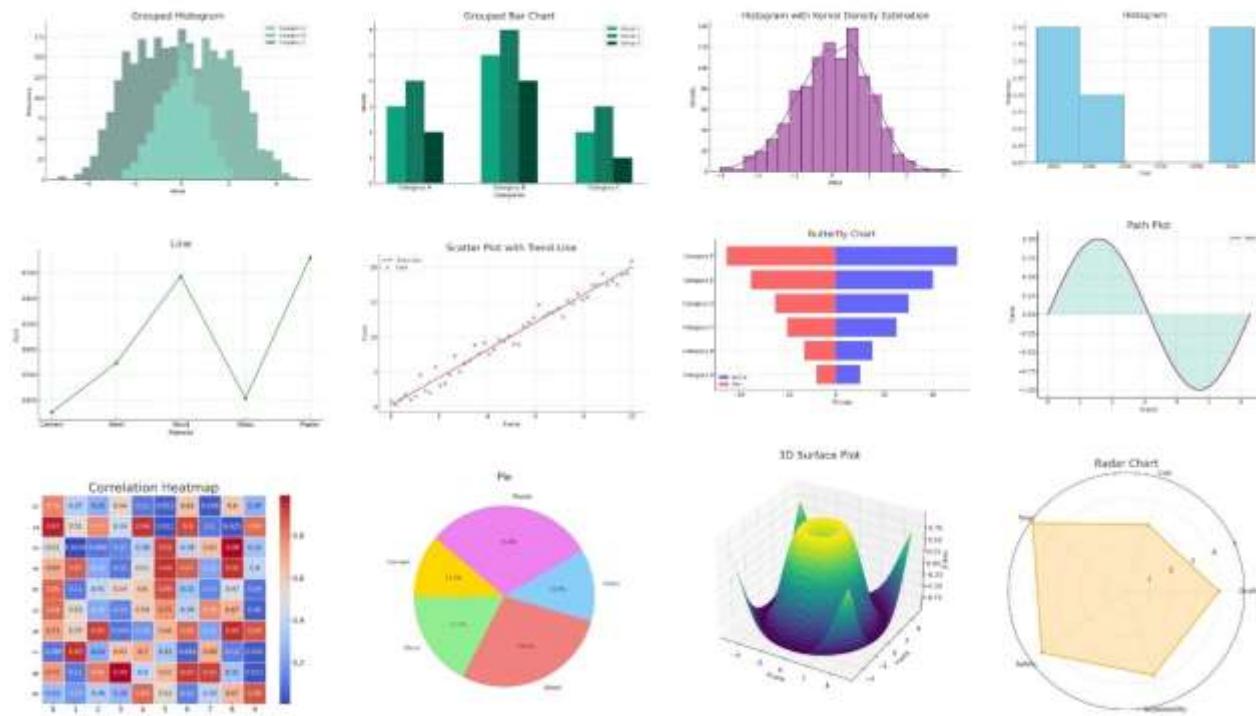
Projektoví manažeři před přijímáním klíčových rozhodnutí častěji používají vizuální znázornění dat než suché a obtížně interpretovatelné údaje z tabulek nebo textových zpráv.

Data bez vizualizace jsou jako stavební materiál nahodile poházený po staveništi: jejich potenciál je nejasný. Teprve když jsou jasně vizualizovány, jako dům z cihel a betonu, je jejich hodnota jasná. Dokud není dům postaven, nelze říci, zda se z hromady materiálu stane malá chatka, luxusní vila nebo mrakodrap.

Společnosti mají k dispozici data z různých systémů (obrázek 1.2-4 až obrázek 2.1-10), finanční transakce a rozsáhlá textová data. Využití těchto dat pro obchodní prospěch je však často náročné. V takových situacích se vizualizace stává důležitým nástrojem pro sdělování významu dat, který pomáhá prezentovat informace ve formátech srozumitelných každému odborníkovi, jako jsou informační panely, grafy a tabulky.

Studie společnosti PwC "What Students Need to Succeed in a Fast Changing Business World" (2015) zdůrazňuje [9], že úspěšné společnosti jdou nad rámec analýzy dat a aktivně využívají interaktivní vizualizační nástroje, jako jsou grafy, infografiky a analytické panely, na podporu rozhodování. Podle zprávy - vizualizace dat pomáhá zákazníkům pochopit příběh, který data vyprávějí, prostřednictvím grafů, diagramů, ovládacích panelů a interaktivních datových modelů.

Proces převodu informací do vizuální grafické podoby, jako jsou tabulky, grafy a diagramy, zlepšuje porozumění a interpretaci dat lidským mozkem (obrázek 7.1-4). To umožňuje projektovým manažerům a analytikům rychleji vyhodnocovat složité scénáře a činit informovaná rozhodnutí na základě vizuálně rozpoznatelných trendů a vzorců, nikoliv intuice.



Obr. 7.1-4 Různé typy vizualizace mají pomocí lidskému mozku lépe pochopit a dát smysl suchým informacím čísel.

Vytváření vizualizací z dat a používání různých volně dostupných vizualizačních knihoven bude podrobněji rozebráno v následující kapitole o ETL -procesech.

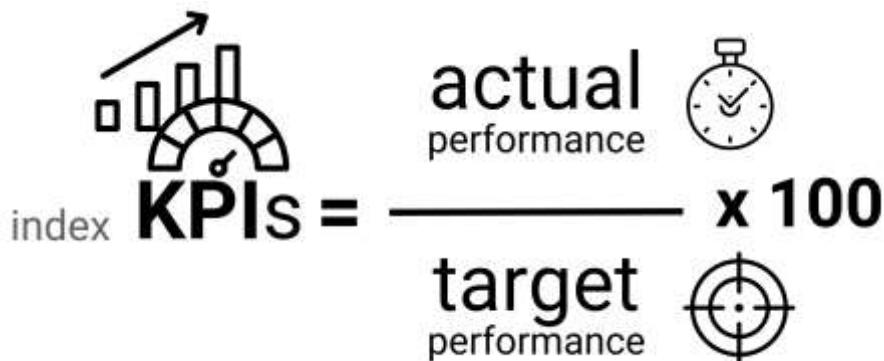
Vizualizace se stává nedílnou součástí práce s daty ve stavebnictví - pomáhá nejen "vidět" data, ale také pochopit jejich význam v kontextu úkolů řízení. Aby však vizualizace byla skutečně užitečná, je třeba předem určit, co přesně by mělo být vizualizováno a které metriky jsou pro hodnocení výkonnosti projektu skutečně důležité. Zde přichází ke slovu výkonnostní metriky, jako jsou KPI a ROI. Bez nich hrozí, že i ty nejkrásnější dashboardy budou jen "informačním šumem".

KPI a ROI

V dnešním stavebnictví hraje správa ukazatelů výkonnosti (KPI a ROI) a jejich vizualizace prostřednictvím zpráv a informačních panelů klíčovou roli při zvyšování produktivity a efektivity řízení projektů.

Stejně jako v každém podnikání je i ve stavebnictví nutné jasně definovat ukazatele, podle kterých se měří úspěch, návratnost investic a výkonnost. Při získávání údajů o různých procesech se organizace, která se řídí daty, musí nejprve naučit identifikovat **klíčové KPI (Key Performance Indicators)** - kvantitativní ukazatele, které odrážejí míru plnění strategických a provozních cílů.

Pro výpočet KPI se obvykle používá vzorec (obrázek 7.1-5), který zahrnuje skutečné a plánované ukazatele. Chcete-li například vypočítat individuální KPI pro projekt, zaměstnance nebo proces, vydělte skutečný výkon plánovaným výkonem a výsledek vynásobte 100%.



Obrázek 7.1-5 KPI se používají k měření úspěšnosti projektu nebo procesu při dosahování klíčových cílů.

Na úrovni pracoviště lze použít podrobnější KPI metriky:

- **Načasování klíčových milníků** (založení, instalace, dokončení) - umožňuje kontrolovat dodržování pracovních plánů.
- **Procento překročení materiálu** - pomáhá řídit nákup a minimalizovat plýtvání.
- **Počet neplánovaných odstávek strojů** - ovlivňuje produktivitu a náklady.

Volba nesprávných metrik může vést k chybným rozhodnutím "co dělat?" (obrázek 2.2-5). Například pokud se společnost zaměří pouze na náklady na metr čtvereční, ale nezohlední náklady na úpravy, mohou úspory na materiálech vést k horší kvalitě a vyšším nákladům v budoucích projektech.

Při stanovování cílů je důležité si ujasnit, co se měří. Nejasná formulace vede k nesprávným závěrům a komplikuje kontrolu. Podívejme se na příklady úspěšných a neúspěšných KPI ve stavebnictví.

Dobré klíčové ukazatele výkonnosti:

- ❑ "Do konce roku snižte podíl rekonstrukčních prací o 10%."
- ❑ "Zvýšení rychlosti montáže fasád o 15% bez snížení kvality do příštího čtvrtletí"
- ❑ "Snížení prostojů strojů o 20% optimalizací pracovních plánů do konce roku"

Tyto metriky jsou jasně měřitelné, s konkrétními hodnotami a časovými rámcemi.

Špatné klíčové ukazatele výkonnosti:

- ❑ "Budeme stavět rychleji" (O kolik rychleji? Co znamená "rychleji"?).
- ❑ "Zlepšíme kvalitu konkrétní práce" (Jak přesně se měří kvalita?)
- ❑ "Zlepšíme interakci s dodavateli na staveništi" (Jaká kritéria prokáží zlepšení?)

Dobrý KPI je takový, který lze měřit a objektivně vyhodnocovat. Ve stavebnictví je to obzvláště důležité, protože bez jasných ukazatelů nelze sledovat výkonnost a dosahovat stabilních výsledků.

Kromě KPI existuje další metrika pro hodnocení efektivnosti investic: **ROI (Return on Investment)** - ukazatel návratnosti investic, který vyjadřuje poměr mezi ziskem a investovanými prostředky. ROI pomáhá posoudit, zda je zavedení nových metod, technologií nebo nástrojů opodstatněné: od digitálních řešení a automatizace (např. obrázek 7.3-2) až po používání nových stavebních materiálů. Tento ukazatel pomáhá činit informovaná rozhodnutí o dalších investicích na základě jejich skutečného dopadu na ziskovost podniku

V kontextu řízení stavebních projektů lze ROI (návratnost investic) použít jako jeden z klíčových ukazatelů výkonnosti (KPI), pokud je cílem společnosti měřit návratnost investic do projektu, technologie nebo zlepšení procesu. Pokud se například zavádí nová technika řízení výstavby, ROI může ukázat, o kolik se díky ní zvýšila ziskovost.

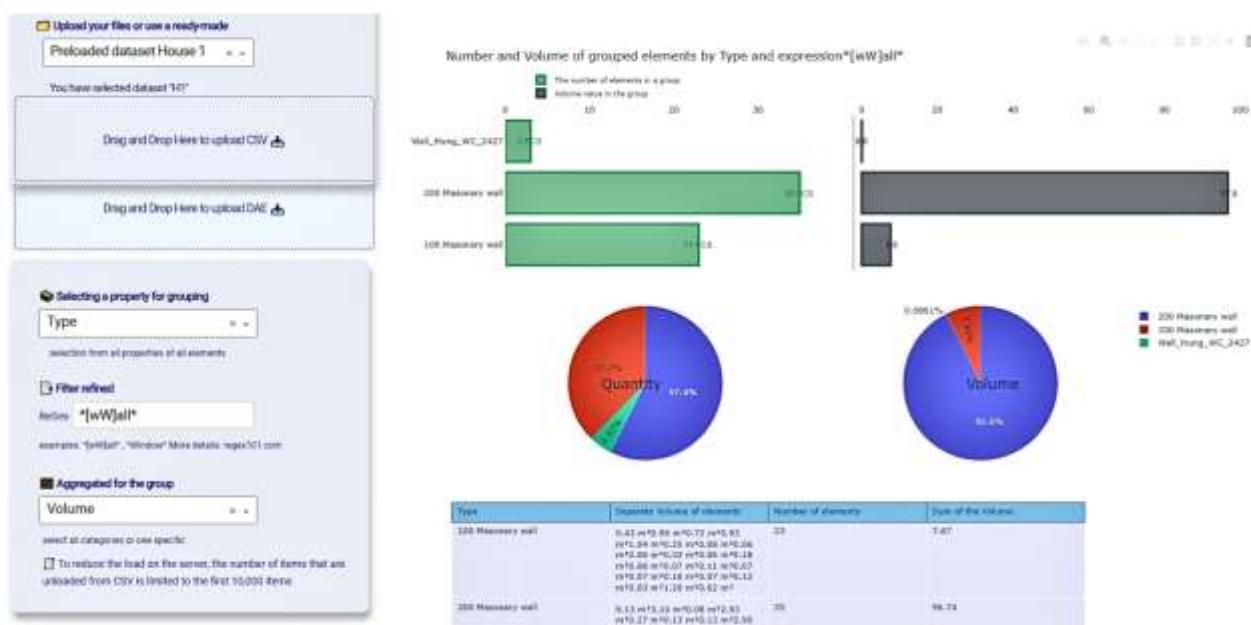
Pravidelné měření klíčových ukazatelů výkonnosti a návratnosti investic na základě údajů získaných z různých zdrojů, jako je spotřeba materiálu, počet hodin práce a náklady, umožňuje vedení projektu efektivně řídit zdroje a přijímat rychlá rozhodnutí. Dlouhodobé ukládání těchto dat umožňuje analyzovat budoucí trendy a optimalizovat procesy.

K vizualizaci klíčových ukazatelů výkonnosti, návratnosti investic a dalších ukazatelů se používají různé grafy a tabulky, které se obvykle spojují do řídicích panelů.

Přístrojové panely a dashboardy: vizualizace metrik pro efektivní řízení

K vizualizaci ukazatelů a metrik se používají různé grafy a diagramy, které se obvykle spojují do datových vitrín a řídicích panelů. Tyto dashboardy poskytují centralizovaný pohled na stav projektu nebo jeho částí a zobrazují klíčové ukazatele (ideálně v reálném čase). Aktuální a průběžně aktualizované dashboardy umožňují týmu rychle reagovat na změny.

Dashboardy jsou nástroje, které vizualizují kvantitativní hodnocení a činí je snadno dostupnými a srozumitelnými pro všechny účastníky projektu.



Obr. 7.1-6 Správa klíčových ukazatelů výkonnosti a jejich vizualizace prostřednictvím informačních panelů je klíčem ke zvýšení produktivity a efektivity projektu.

Zde je několik příkladů oblíbených nástrojů, ve kterých můžete vytvářet řídící panely:

- **Power BI** je nástroj od společnosti Microsoft pro vytváření interaktivních sestav a řídících panelů.
- **Tableau a Google Data Studio** jsou výkonné nástroje pro vizualizaci dat a vytváření řídících panelů bez nutnosti psát kód.
- **Plotly** (obrázek 7.1-6, obrázek 7.2-12) je knihovna pro vytváření interaktivních grafů a Dash je framework pro vytváření webových aplikací pro analýzu dat. Jejich kombinací lze vytvářet interaktivní informační panely.
- **Četné knihovny Pythonu** (obrázek 7.2-9 - obrázek 7.2-11) - Python má mnoho open source a bezplatných knihoven pro vizualizaci dat, například Matplotlib, Seaborn, Plotly, Bokeh a další. Ty lze použít k vytváření grafů a jejich integraci do webové aplikace pomocí frameworků, jako je Flask nebo Django.
- **Knihovny JavaScriptu**: umožňuje vytvářet interaktivní informační panely pomocí knihoven JavaScriptu s otevřeným zdrojovým kódem, jako je D3.js nebo Chart.js, a integrovat je do webové aplikace.

K vyhodnocování klíčových ukazatelů výkonnosti a vytváření informačních panelů potřebujete aktuální data a jasný harmonogram sběru a analýzy informací.

Obecně platí, že KPI, ROI a informační panely ve stavebnictví tvoří základ analytického přístupu k řízení projektů. Pomáhají nejen sledovat a vyhodnocovat aktuální stav, ale také poskytují cenné poznatky pro

budoucí plánovací a optimalizační procesy - procesy, které přímo závisí na interpretaci dat a kladení správných a včasných otázek.

Analýza dat a umění klást otázky

Interpretace dat je závěrečnou fází analýzy, kdy informace dávají smysl a začínají "mluvit". Zde se formulují odpovědi na klíčové otázky: "Co dělat?" a "Jak dělat?" (obrázek 2.2-5). Tato fáze umožňuje shrnout výsledky, identifikovat vzorce, stanovit příčinné vztahy a vyvodit závěry na základě vizualizace a statistické analýzy.

Možná není daleko doba, kdy si uvědomí, že k tomu, aby se stal plně efektivním občanem jednoho z velkých nových komplexních světových států, které se nyní rozvíjejí, je stejně nutné umět počítat, myslet v termínech průměrů, maxim a minim, jako je dnes nutné umět číst a psát [133].

- Samuel S. Wilkes, citát z prezidentského projevu před Americkou statistickou asociací v roce 1951.

Podle zprávy "Analýza dat a umělá inteligence při realizaci vládních projektů" (2024), kterou zveřejnila vláda Spojeného království [83], může implementace analytiky dat a umělé inteligence (AI) významně zlepšit procesy řízení projektů, zvýšit přesnost předvídání času a nákladů a také snížit rizika a nejistotu. Dokument zdůrazňuje, že veřejné organizace, které využívají pokročilé analytické nástroje, dosahují vyšší výkonnosti v infrastrukturních iniciativách.

Moderní stavební podnikání ve vysoce konkurenčním prostředí čtvrté průmyslové revoluce s nízkými maržemi lze přirovat k vojenským operacím. Přežití a úspěch podniku zde závisí na rychlosti získávání zdrojů a kvalitních informací - a tedy na včasném a informovaném rozhodování (obr. 7.1-7).

Pokud je vizualizace dat "zpravodajstvím", které poskytuje přehled, pak analýza dat je "municí" potřebnou k akci. Odpovídá na otázky: "Co dělat?" a "Jak to dělat?" a tvoří základ pro získání konkurenční výhody na trhu.

Analytika mění různorodá data na strukturované a smysluplné informace, na jejichž základě se lze rozhodovat.

Úkolem analytiků a manažerů není pouze interpretovat informace, ale nabízet informovaná rozhodnutí, identifikovat trendy, určovat vztahy mezi různými typy dat a kategorizovat je v souladu s cíli a specifiky projektu. Pomocí vizualizačních nástrojů a metod statistické analýzy proměňují data ve strategický přínos pro podnik.



Obr. 7.1-7 Je to analýza dat, která nakonec ze získaných informací udělá zdroj pro rozhodování.

Aby bylo možné v analytickém procesu přijímat skutečně informovaná rozhodnutí, je nutné naučit se správně formulovat otázky, které jsou datům kladené. Kvalita těchto otázek přímo ovlivňuje hloubku získaných poznatků a v důsledku toho i kvalitu manažerských rozhodnutí.

Minulost existuje pouze do té míry, do jaké je přítomna v dnešních záznamech. A to, co tyto záznamy představují, je dáno otázkami, které si klademe. Neexistuje žádná jiná historie než tato [134].

- John Archibald Wheeler, fyzik 1982

Umění klást hluboké otázky a kriticky myslit je při práci s daty nezbytnou dovedností. Většina lidí má tendenci klást jednoduché, povrchní otázky, jejichž zodpovězení nevyžaduje velké úsilí. Skutečná analýza však začíná smysluplnými a promyšlenými otázkami, které mohou odhalit skryté vztahy a příčinné souvislosti v informacích, jež mohou být skryty za několika vrstvami uvažování.

Podle studie "Data-Driven Transformation: Accelerating at Scale Now" (BCG, 2017) [135] vyžaduje úspěšná digitální transformace investice do analytických schopností, programů řízení změn a sladění obchodních cílů s IT iniciativami. Společnosti, které vytvářejí kulturu založenou na datech, by měly investovat do kapacit datové analytiky a zahájit programy řízení změn, aby si vštípily nové myšlení, chování a způsoby práce.

Bez investic do rozvoje analytické kultury, zdokonalování datových nástrojů a školení odborníků budou společnosti i nadále riskovat, že budou přijímat rozhodnutí na základě zastaralých nebo neúplných informací - nebo se budou spoléhat na subjektivní názory manažerů HiPPO (obrázek 2.1-9).

Uvědomění si významu a potřeby neustále aktualizovat analytické a dashboardy nevyhnutelně vede management k pochopení významu automatizace analytických procesů. Automatizace zvyšuje rychlosť rozhodování, snižuje vliv lidského faktoru a zajišťuje relevanci dat. S exponenciálním růstem objemu informací se rychlosť stává nejen konkurenční výhodou, ale i klíčovým faktorem udržitelného úspěchu.

Automatizace procesů analýzy a zpracování dat obecně je neoddělitelně spjata s tématem ETL (Extract, Transform, Load). Stejně jako v procesu automatizace musíme data transformovat, v procesu

ETL jsou data extrahována z různých zdrojů, transformována podle potřebných požadavků a načtena do cílových systémů pro další použití.



KAPITOLA 7.2.

TOK DAT BEZ RUČNÍ PRÁCE: PROČ JE POTŘEBA ETL

Automatizace ETL: snížení nákladů a zrychlení zpracování dat

Když klíčové ukazatele výkonnosti (KPI) přestanou růst navzdory nárůstu objemu dat a velikosti týmu, vedení společnosti si nevyhnutelně uvědomí potřebu automatizace procesů. Dříve či později se toto uvědomění stane podnětem k zahájení komplexní automatizace, jejímž hlavním cílem je snížit složitost procesů, zrychlit zpracování a snížit závislost na lidském faktoru.

Podle studie společnosti McKinsey "How to Build a Data Architecture to Drive Innovation - Today and Tomorrow" (2022) [136] získávají společnosti využívající architektury proudových dat významnou výhodu, protože mohou analyzovat informace v reálném čase. Streamovací technologie umožňují přímou analýzu zpráv v reálném čase a aplikaci prediktivní údržby ve výrobě prostřednictvím analýzy dat ze senzorů v reálném čase.

Zjednodušení procesů je automatizace, kdy jsou tradiční manuální funkce nahrazeny algoritmy a systémy.

Otázka automatizace, respektive "minimalizace role člověka při zpracování dat", je pro každou společnost nezvratným a velmi citlivým procesem. Specialisté v jakémkoli profesním oboru často váhají, zda mají své metody a jemnosti práce plně odhalit kolegům optimalizátorům, protože si uvědomují riziko ztráty zaměstnání v rychle se vyvíjejícím technologickém prostředí.

Pokud si chcete udělat nepřátele, snažte se věci změnit [137].

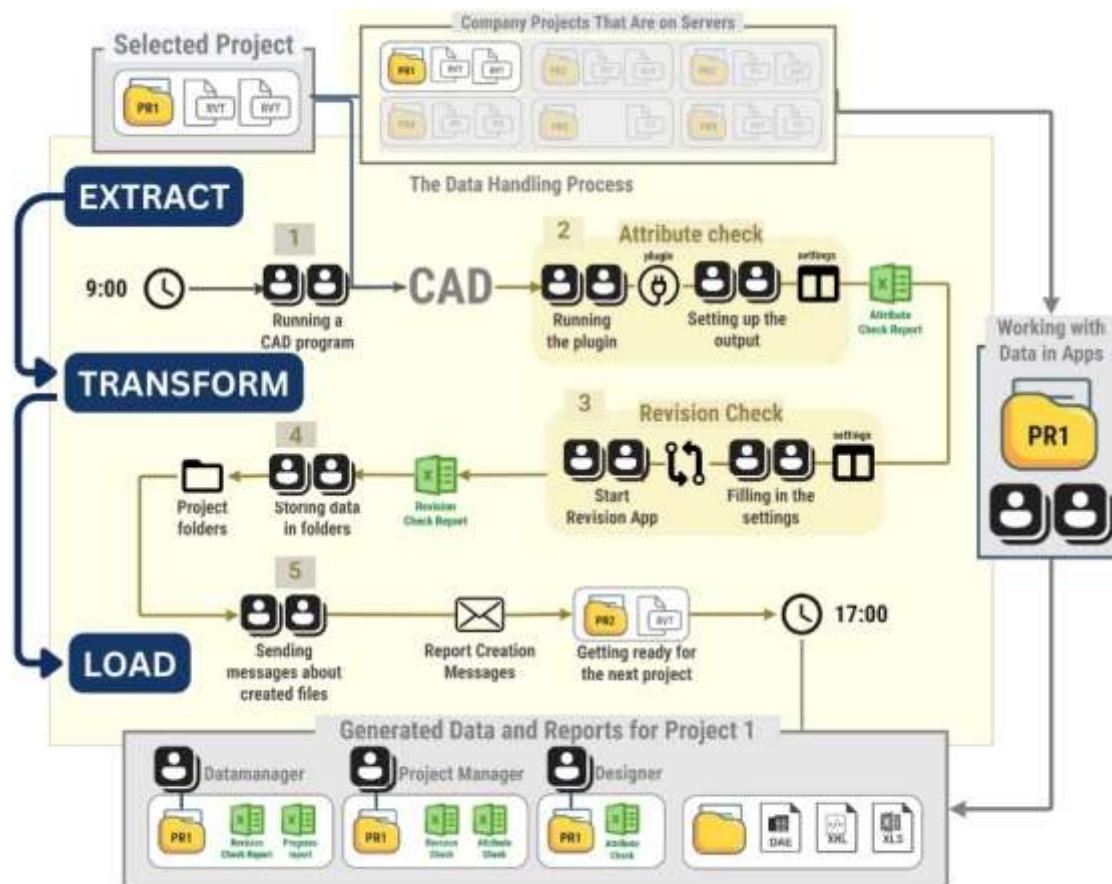
- Woodrow Wilson, projev na kongresu obchodníků, Detroit, 1916

Navzdory zřejmým výhodám automatizace má mnoho společností ve své každodenní praxi stále vysoký podíl ruční práce, zejména v oblasti technických dat. Pro ilustraci současné situace se podívejme na typický příklad sekvenčního zpracování dat v rámci těchto procesů.

Ruční zpracování dat lze ilustrovat na příkladu interakce s informacemi získanými z databází CAD. Tradiční zpracování dat ("ruční" ETL -proces) v odděleních CAD (BIM) pro tvorbu atributových tabulek nebo tvorbu dokumentace na základě dat o návrhu probíhá v následujícím pořadí (obr. 7.2-1):

1. **Ruční extrakce (Extract):** uživatel otevře projekt ručně - spuštěním aplikace CAD (BIM) (obr. 7.2-1 krok 1).
2. **Ověření:** Další krok obvykle zahrnuje ruční spuštění několika zásuvných modulů nebo pomocných aplikací, které připraví data a posoudí jejich kvalitu (obr. 7.2-1 krok 2-3).
3. **Ruční transformace (Transformace):** Po přípravě začíná zpracování dat, které vyžaduje ruční obsluhu různých softwarových nástrojů, v nichž se data připravují k odeslání (obr. 7.2-1 krok

- 4).
4. Ruční **nahrávání (Load)**: ruční nahrávání převedených dat do externích systémů, datových formátů a dokumentů (obr. 7.2-1 krok 5).



Obr. 7.2-1 Tradiční ruční zpracování ETL je omezeno přání a fyzickými možnostmi jednotlivých techniků.

Takový pracovní postup je příkladem klasického procesu ETL - extrakce, transformace a načítání (ETL). Na rozdíl od jiných průmyslových odvětví, kde jsou automatické ETL pipelines již dlouho standardem, ve stavebnictví stále převažuje manuální práce, která zpomaluje procesy a zvyšuje náklady

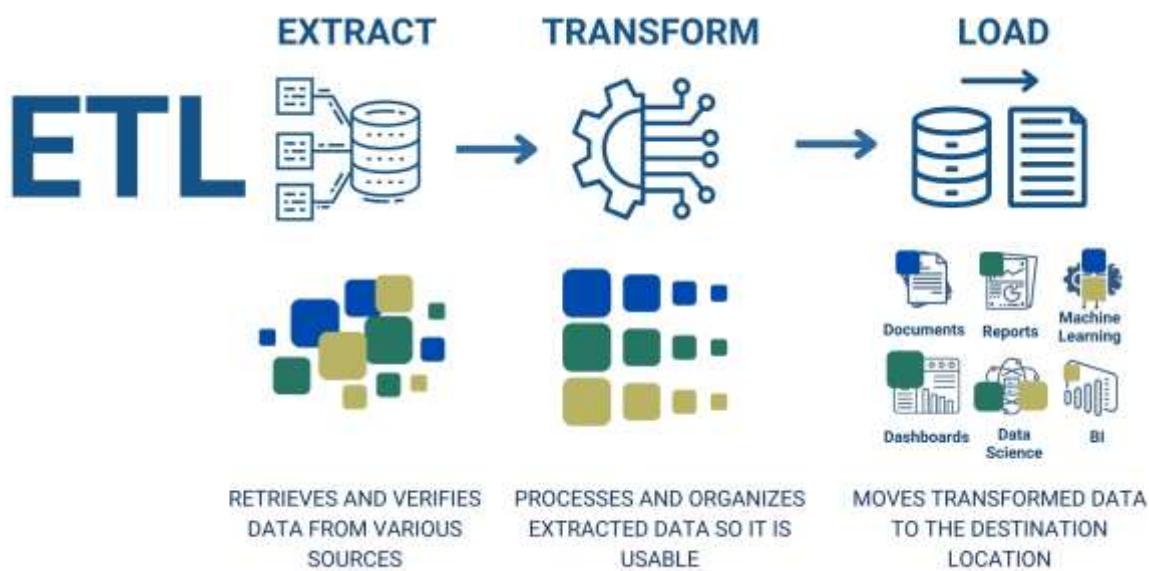
ETL (Extract, Transform, Load) je proces extrakce dat z různých zdrojů, jejich transformace do požadovaného formátu a načtení do cílového systému pro další analýzu a použití.

ETL je proces, který označuje tři klíčové součásti zpracování dat: Extract, Transform a Load (obrázek 7.2-2):

- Extrakce - extrakce dat z různých zdrojů (soubory, databáze, API).

- **Transformace** - čištění dat, agregace, normalizace a logické zpracování.
- **Load** - načítání strukturovaných informací do datového skladu, sestav nebo systému BI.

Dříve v knize jsme se konceptu ETL dotkli jen příležitostně: při převodu nestrukturovaného naskenovaného dokumentu do strukturovaného tabulkového formátu (obr. 4.1-1), v souvislosti s formalizací požadavků pro systematizaci vnímání životních i obchodních procesů (obr. 4.4-20) a při automatizaci validace a zpracování dat z CAD řešení. Podívejme se nyní na ETL podrobněji v kontextu typických pracovních postupů.



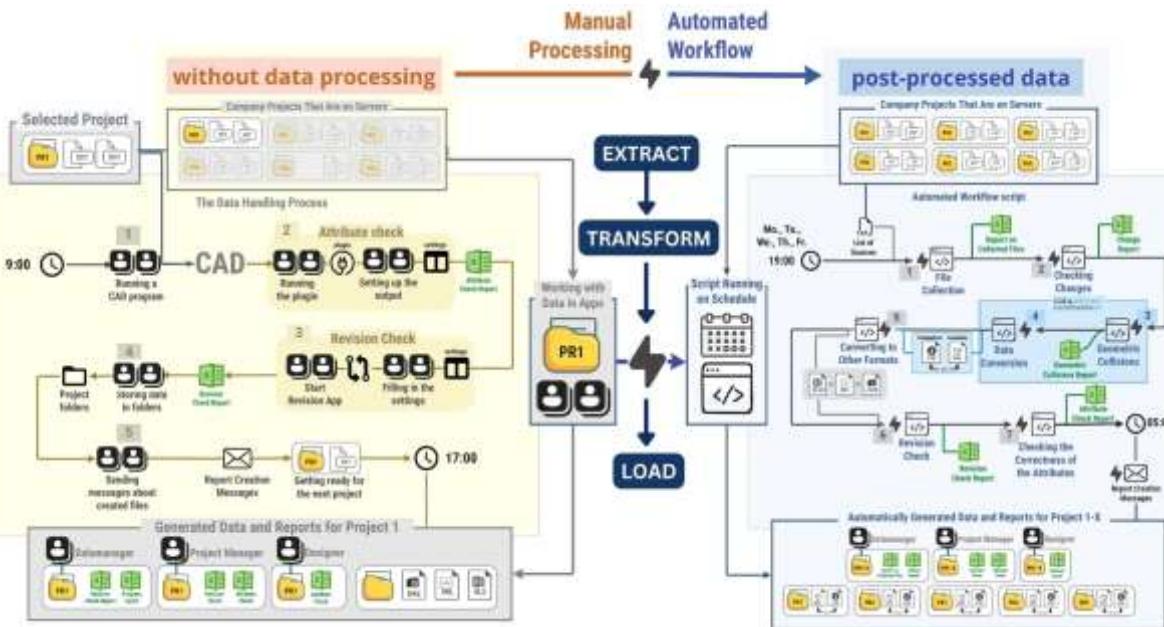
Obrázek 7.2-2 ETL automatizuje opakující se úlohy zpracování dat.

Manuální nebo poloautomatizovaný ETL -proces předpokládá manažera nebo technika, který řídí všechny kroky ručně - od sběru dat až po generování sestav. Takový proces zabere značný čas, zejména pokud je pracovní doba omezená (např. od 9:00 do 17:00).

Podniky se často snaží vyřešit problém nízké efektivity a pomalé rychlosti nákupem modulárních integrovaných řešení (ERP, PMIS, CPM, CAFM, atd.), která jsou pak dále rozvíjena externími dodavateli a konzultanty. Tito dodavatelé a externí vývojáři se však často stávají kritickým místem závislosti: jejich technická omezení přímo ovlivňují výkonnost celého systému a podniku jako celku, jak bylo podrobně popsáno v předchozích kapitolách o proprietárních systémech a formátech. Problémy vzniklé roztríštěností a závislostmi byly podrobně rozebrány v kapitole "Jak se stavební firmy topí v datovém chaosu".

Pokud společnost není připravena implementovat velkou modulární platformu od některého z dodavatelů, začne hledat alternativní způsoby automatizace. Jednou z nich je vývoj vlastních modulárních otevřených ETL -konvejerů, kde jsou jednotlivé fáze (extrakce, transformace, validace, načítání) implementovány jako skripty spouštěné podle plánu.

V automatizované verzi téhož pracovního postupu ETL (obr. 7.2-1) vypadá pracovní postup jako modulární kód, který začíná zpracováním dat a jejich převodem do otevřené strukturované podoby. Jakmile jsou strukturovaná data přijata, automaticky se podle plánu spouštějí různé skripty nebo moduly, které kontrolují změny, transformují a odesírají zprávy (obr. 7.2-3).



Obr. 7.2-3 Vlevo je ruční obrábění, vpravo automatický proces, který na rozdíl od tradičního ručního obrábění není omezen možnostmi uživatele.

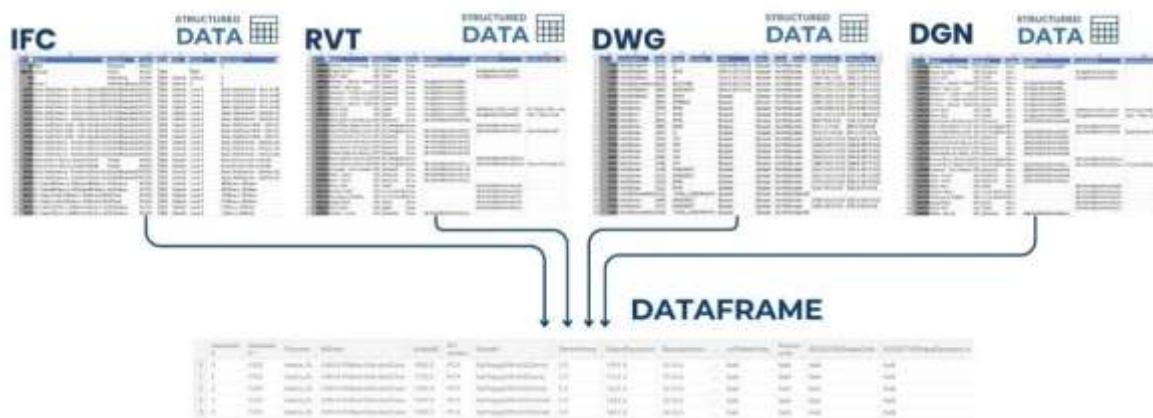
V automatizovaném pracovním postupu se zpracování dat zjednoduší pomocí předzpracování dat ET(L): strukturování a sjednocení.

Při tradičních metodách zpracování pracují odborníci s daty "tak, jak jsou" - jak jsou získávána ze systémů nebo softwaru. V automatizovaných procesech naopak data často nejprve procházejí ETL -payplane, kde jsou uvedena do konzistentní struktury a formátu vhodného pro další použití a analýzu.

Uveďme si praktický příklad ETL, který demonstruje proces validace datové tabulky popsáνý v kapitole "Validace dat a výsledky validace" (obrázek 4.4-13). K tomu použijeme knihovnu Pandas ve spojení s LLM pro automatizované procesy analýzy a zpracování dat.

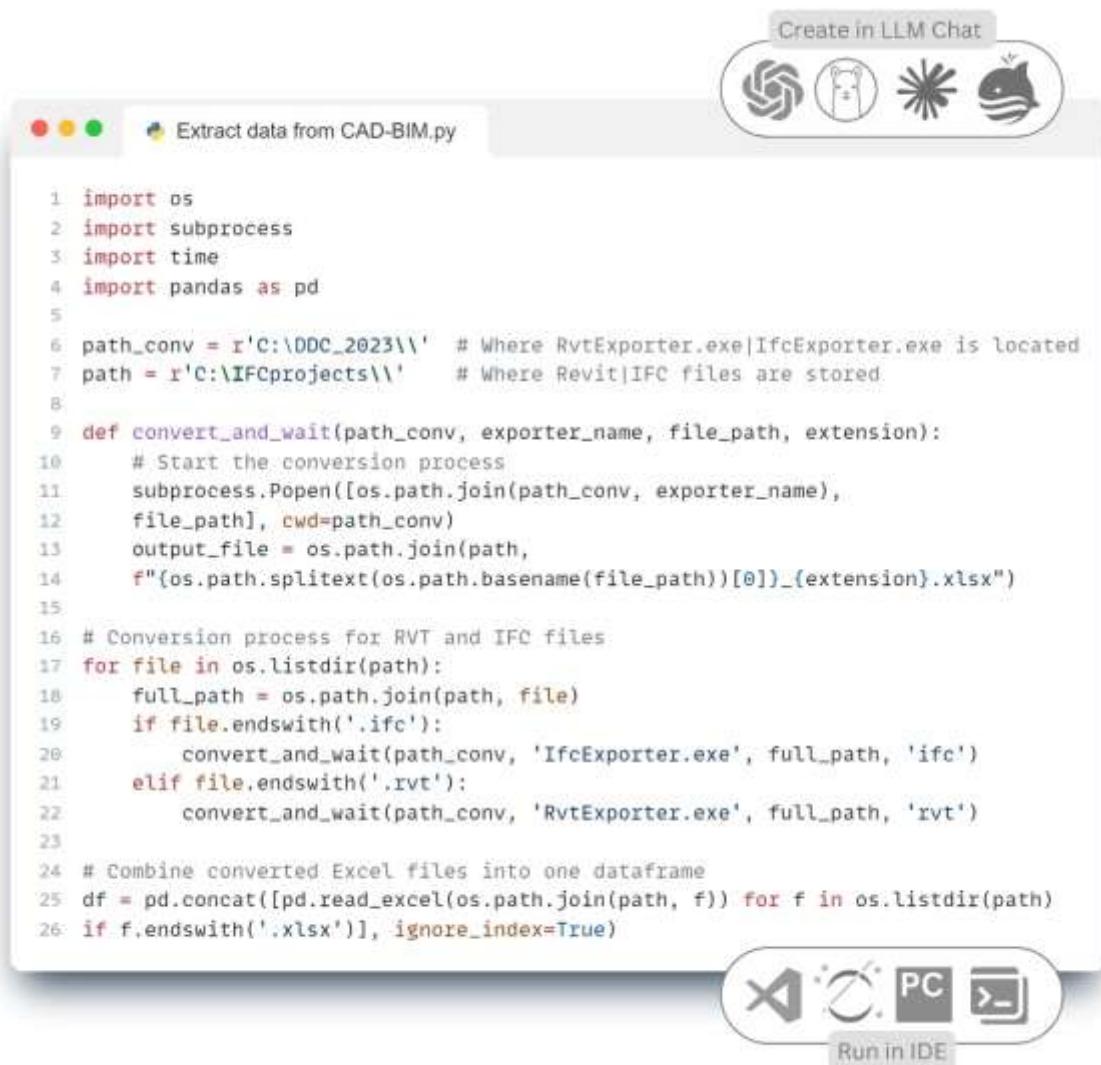
ETL Výpis: sběr dat

První fáze procesu ETL - Extract) - začíná psaním kódu, který shromažďuje sady dat, jež mají být dále kontrolovány a zpracovány. Za tímto účelem prohledáme všechny složky produkčního serveru, shromázdíme dokumenty určitého formátu a obsahu a poté je převedeme do strukturované podoby. Tento proces je podrobně popsán v kapitolách "Převod nestrukturovaných a textových dat do strukturované podoby" a "Převod dat CAD (BIM) do strukturované podoby" (obr. 4.1-1 - obr. 4.1-12).



Obr. 7.2-4 Převést data CAD (BIM) do jednoho velkého datového rámce, který bude obsahovat všechny části projektu.

Jako ilustrativní příklad použijeme krok Extract načtení dat a získáme tabulkou všech projektů CAD-(BIM-) (obr. 7.2-4) používá konvertory s podporou reverzního inženýrství [138] formáty RVT a IFC k získání strukturovaných tabulek ze všech projektů a jejich spojení do jedné velké tabulky DataFrame.



```

1 import os
2 import subprocess
3 import time
4 import pandas as pd
5
6 path_conv = r'C:\DDC_2023\' # Where RvtExporter.exe|IfcExporter.exe is located
7 path = r'C:\IFCprojects\' # Where Revit|IFC files are stored
8
9 def convert_and_wait(path_conv, exporter_name, file_path, extension):
10    # Start the conversion process
11    subprocess.Popen([os.path.join(path_conv, exporter_name),
12                      file_path], cwd=path_conv)
13    output_file = os.path.join(path,
14                               f"{os.path.splitext(os.path.basename(file_path))[0]}_{extension}.xlsx")
15
16    # Conversion process for RVT and IFC files
17    for file in os.listdir(path):
18        full_path = os.path.join(path, file)
19        if file.endswith('.ifc'):
20            convert_and_wait(path_conv, 'IfcExporter.exe', full_path, 'ifc')
21        elif file.endswith('.rvt'):
22            convert_and_wait(path_conv, 'RvtExporter.exe', full_path, 'rvt')
23
24    # Combine converted Excel files into one dataframe
25 df = pd.concat([pd.read_excel(os.path.join(path, f)) for f in os.listdir(path)
26 if f.endswith('.xlsx')], ignore_index=True)

```

Run in IDE

Obr. 7.2-5 Převod pomocí kódu Python a nástroje SDK pro reverzní inženýrství pro soubory RVT a IFC do jednoho velkého strukturovaného (df) DataFrame.

Pandas DataFrame může načítat data z různých zdrojů, včetně textových souborů CSV, tabulek Excel, souborů JSON - a XML -, formátů pro ukládání velkých objemů dat, jako jsou Parquet a HDF5, a z databází MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Oracle a dalších. Kromě toho Pandas podporuje načítání dat z rozhraní API, webových stránek, cloudových služeb a úložných systémů, jako jsou Google BigQuery, Amazon Redshift a Snowflake.

- 💡 Chcete-li napsat kód pro připojení a sběr informací z databází, pošlete podobný textový požadavek na chat LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN nebo jakýkoli jiný):

Napište prosím příklad připojení k MySQL a převodu dat do .

💡 Odpověď společnosti LLM :

```

1 import pandas as pd
2 import mysql.connector
3
4 # Using a context manager to ensure
5 # that the connection is closed automatically
6 with mysql.connector.connect(host='hostname', database='dbname',
7     user='username', password='password') as conn:
8     df = pd.read_sql('SELECT * FROM tablename LIMIT 10', conn)

```

Create in LLM Chat

Run in IDE

Obrázek 7.2-6 Příklad připojení přes Python k databázi MySQL a import dat z databáze MySQL do DataFrame .

Výsledný kód (Obr. 7.2-5, Obr. 7.2-6) lze spustit v některém z populárních IDE (integrovaných vývojových prostředí), která jsme zmínili výše, v offline režimu: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse s pluginem PyDev, Thonny, Wing IDE, IntelliJ IDEA s pluginem Python, JupyterLab nebo populární online nástroje: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Načtením dat multiformátu do proměnné "df" (obrázek 7.2-5 - řádek 25; obrázek 7.2-6 - řádek 8) jsme data převedli do formátu Pandas DataFrame, jedné z nejoblíbenějších struktur pro zpracování dat, což je dvouozměrná tabulka s řádky a sloupcí. O dalších formátech pro ukládání dat používaných v ETL - Pipeline, jako jsou Parquet, Apache ORC, JSON, Feather, HDF5, a také o moderních datových skladech si povíme více v kapitole "Ukládání a správa dat ve stavebnictví" (obrázek 8.1-2).

Po fázi extrakce a strukturování dat (Extract) vznikne jediné pole informací (obr. 7.2-5, obr. 7.2-6), připravené k dalšímu zpracování. Před načtením těchto dat do cílových systémů nebo jejich použitím pro analýzu je však nutné zajistit jejich kvalitu, integritu a soulad se stanovenými požadavky. V této fázi dochází k transformaci dat (Transform) - klíčovému kroku pro zajištění spolehlivosti následných závěrů a rozhodnutí.

ETL Transformace: použití validačních a transformačních pravidel.

V kroku Transformace se data zpracovávají a transformují. Tento proces může zahrnovat kontrolu

správnosti, normalizaci, doplnění chybějících hodnot a validaci pomocí automatizovaných nástrojů

Podle studie PwC "Data-Driven. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9] se moderní auditorské firmy odklánějí od náhodného ověřování dat a směřují k analýze obrovského množství informací pomocí automatizovaných nástrojů. Tento přístup umožňuje nejen identifikovat nesrovnanosti ve výkaznictví, ale také nabízet doporučení pro optimalizaci podnikových procesů.

Ve stavebnictví lze podobné metody použít například pro automatickou validaci dat návrhu, kontrolu kvality stavby a hodnocení výkonnosti dodavatele. Jedním z nástrojů pro automatizaci a urychlení zpracování dat je použití regulérních výrazů (RegEx) ve fázi transformace dat (Transform) procesu ETL. RegEx umožňuje efektivně ověřovat datové řetězce, odhalovat nekonzistence a zajišťovat integritu informací s minimální spotřebou zdrojů. Podrobnější informace o RegEx (obr. 4.4-7) byly probrány v kapitole "Převod požadavků do strukturované podoby".

Uvedeme si praktický příklad: v systému správy objektů majetku (RPM) správce nastavuje požadavky na klíčové atributy objektů (obr. 7.2-7). Ve fázi transformace je třeba ověřit následující parametry:

- ověření formátu identifikátoru objektu (atribut "ID")
- kontrola hodnot náhradní záruční doby (atribut "Záruční doba")
- ověření cyklu výměny prvků (atribut "Požadavky na údržbu").

ID	Element	Warranty Period	Replacement Cycle	Maintenance Requirements
W-NEW	Window	-	20 years	Annual Inspection
W-OLD1	Window	8 years	15 years	Biannual Inspection
W-OLD2	Window	8 years	15 years	Biannual Inspection
D-122	Door	15 years	25 years	Biennial Varnishing

Obrázek 7.2-7 Ověřování kvality začíná stanovením požadavků na atributy a jejich mezních hodnot.

Pro nastavení hraničních hodnot pro ověřování parametrů například předpokládejme, že z našich zkušeností víme, že platné hodnoty pro atribut "ID" mohou obsahovat pouze řetězcové hodnoty "W-NEW", "W-OLD1" nebo "D-122" nebo podobné hodnoty, kde prvním znakem je písmeno následované pomlčkou a pak tři abecední znaky "NEW", "OLD" nebo libovolné třímístné číslo (obrázek 7.2-7). K ověření těchto identifikátorů lze použít následující regulérní výraz (RegEx):

```
^W-NEW$ | ^W-OLD[0-9]+$ | ^D-1[0-9]{2}$
```

Tato šablona zajistí, aby všechny identifikátory v datech odpovídaly zadaným kritériím. Pokud některá hodnota kontrolou neprojde, systém zaznamená chybu. Chcete-li vytvořit -kód jazyka Python pro transformaci dat a použít výsledná data k vytvoření tabulky výsledků, stačí formulovat dotaz v chatu LLM.

⌚ Textová žádost na adresu LLM:

Napište kód pro ověření sloupců DataFrame pomocí regulárních výrazů, který kontroluje identifikátory ve formátu "W-NEW" nebo "W-OLD" pomocí RegEx, energetickou účinnost s písmeny "A" až "G", záruční dobu a cyklus výměny s číselnými hodnotami v letech. ↴

💡 Odpověď společnosti LLM :

```

1 # RegEx
2 import re
3
4 # Define regex patterns for each column
5 patterns = {
6     'Window_ID': r'^(W-NEW|W-OLD\d{1,2}\D{1}\d{1}\d{2})$',
7     'Energy_Efficiency': r'^[A-G]\w*$',
8     'Warranty_Period': r'^\d+\syears$',
9     'Replacement_Cycle': r'^\d+\syears$'
10 }
11
12 # Define a verification function using regex
13 def verify_data(column, pattern):
14     # Apply the regex pattern to the series and return the boolean results
15     return column.apply(lambda x: bool(re.match(pattern, str(x))))
16
17 # Apply the verification function to the DataFrame for each column with its respective pattern
18 for column, pattern in patterns.items():
19     df['verified_' + column] = verify_data(df[column], pattern)
20

```

Obrázek 7.2-8 Kód automatizuje proces validace použitím šablon RegEx na sloupce parametrů datového rámu .

Výše uvedený, automaticky generovaný kód jazyka Python (obrázek 7.2-8), používá knihovnu "re" (regulární výrazy RegEx) k definici funkce, která kontroluje každý atribut datové položky v DataFrame. Pro každý zadaný sloupec (atribut) použije funkce vzor RegEx, aby zkontrolovala, zda každý záznam odpovídá očekávanému formátu, a přidá výsledky jako nové hodnoty (False/True) do nového atributu-sloupce DataFrame.

Tato automatizovaná validace zajišťuje formální shodu údajů a může být použita jako součást systému kontroly kvality ve fázi transformace.

Po úspěšném dokončení kroku transformace a kontroly kvality jsou data připravena k odeslání do cílových systémů. Transformovaná a ověřená data lze pro další použití nahrát do CSV, JSON, Excel, databází a dalších formátů. V závislosti na úloze lze výsledky také prezentovat v sestavách, grafech nebo analytických panelech.

ETL Load: Vizualizace výsledků v tabulkách a grafech.

Po dokončení fáze Transformace, kdy jsou data strukturována a ověřena, následuje závěrečná fáze Načíst, kdy lze data načíst do cílového systému i vizualizovat pro účely analýzy. Vizuální prezentace dat umožňuje rychle identifikovat odchylinky, analyzovat rozdělení a sdělit klíčové závěry všem účastníkům projektu, včetně těch bez technického vzdělání.

Místo prezentace informací v podobě tabulek a čísel můžeme používat infografiky, grafy a řídicí panely (dashboardy). Jedním z nejběžnějších a nejflexibilnějších nástrojů pro vizualizaci strukturovaných dat v jazyce Python je knihovna Matplotlib (obrázek 7.2-9, obrázek 7.2-10). Umožňuje vytvářet statické, animované a interaktivní grafy a podporuje širokou škálu typů grafů.

- ☒ Pro vizualizaci výsledků kontroly atributů ze systému RPM (obrázek 7.2-7), můžete použít následující dotaz na jazykový model:

Napište kód pro vizualizaci dat DataFrame, výše (obrázek 7.2-7), s histogramem pro výsledky, který ukáže chybovost atributu. ↵

- ❸ Odpověď LLM ve formě kódu a připravená vizualizace výsledků provádění kódu přímo v chatu LLM :

```

Create in LLM Chat
Pipeline.py
1 # Re-importing necessary libraries for visualization
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4
5 # Visualization 1: Bar Chart
6 plt.figure(figsize=(10, 6))
7 df_visual.plot(kind='bar', stacked=True, color=['green', 'red'])
8 plt.title('Data Verification Summary - Bar Chart')
9 plt.xlabel('Data Categories')
10 plt.ylabel('Count')
11 plt.xticks(rotation=45)
12 plt.tight_layout()
13 plt.show()

```

Run in IDE

Data Verification Summary - Bar Chart

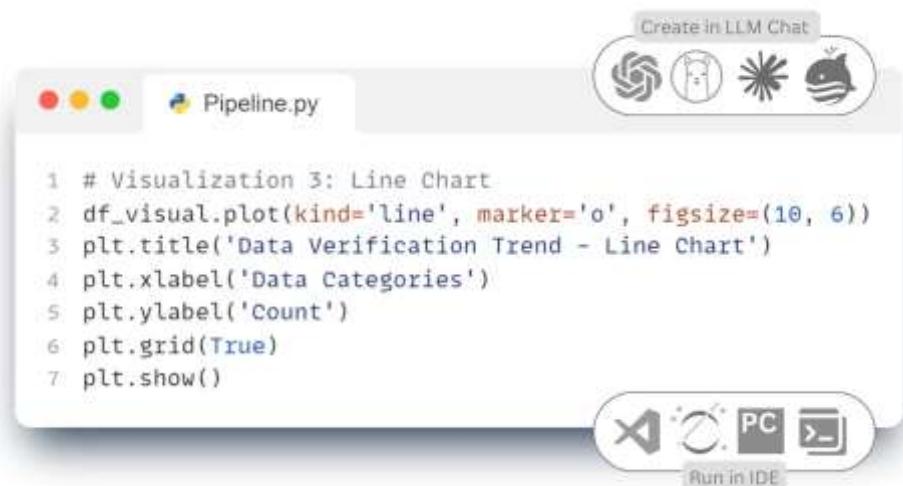
Data Category	Passed	Failed
Window_ID	4	1
Energy_Efficiency	3	2
Warranty_Period	4	1
Replacement_Cycle	4	1

Obr. 7.2-9 Vizualizace výsledků kroku Transformace kontroly hodnot atributů ze systému RPM (obr. 7.2-7) jako histogram v kroku Načíst .

- 2 Existuje mnoho volně dostupných vizualizačních knihoven s otevřeným zdrojovým kódem, které umožňují prezentovat strukturovaná data v různých formátech. Pokračujme ve vizualizaci výsledků pomocí jiného typu grafu s následující výzvou v chatu:

Vykreslete stejná data jako graf přímek ↪

- 2 Odpověď společnosti LLM :

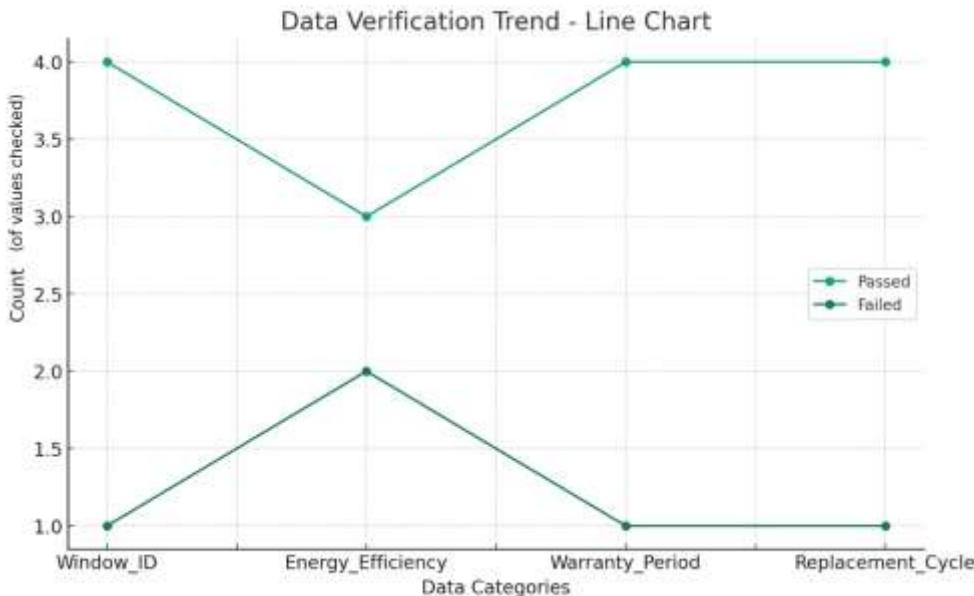


```

1 # Visualization 3: Line Chart
2 df_visual.plot(kind='line', marker='o', figsize=(10, 6))
3 plt.title('Data Verification Trend - Line Chart')
4 plt.xlabel('Data Categories')
5 plt.ylabel('Count')
6 plt.grid(True)
7 plt.show()

```

The screenshot shows a Jupyter Notebook cell with the code above. Below the cell is a line chart titled "Data Verification Trend - Line Chart". The x-axis is labeled "Data Categories" and has four points: "Window_ID", "Energy_Efficiency", "Warranty_Period", and "Replacement_Cycle". The y-axis is labeled "Count (of values checked)" and ranges from 1.0 to 4.0. There are two data series: "Passed" (green line with circles) and "Failed" (red line with circles). The "Passed" series has values [4.0, 3.0, 4.0, 4.0] and the "Failed" series has values [1.0, 2.0, 1.0, 1.0].



Obr. 7.2-10 Vizualizace validačních dat (obr. 7.2-8) jako čárový diagram získaný pomocí knihovny Matplotlib .

Existuje mnoho bezplatných vizualizačních knihoven s otevřeným zdrojovým kódem, např.:

- Seaborn - pro statistické grafy (obr. 7.2-11)
- Plotly - pro interaktivní webové vizualizace (obr. 7.2-12, obr. 7.1-6).
- Altair - pro deklarativní vizualizaci
- Dash nebo Streamlit - pro vytvoření kompletních ovládacích panelů.

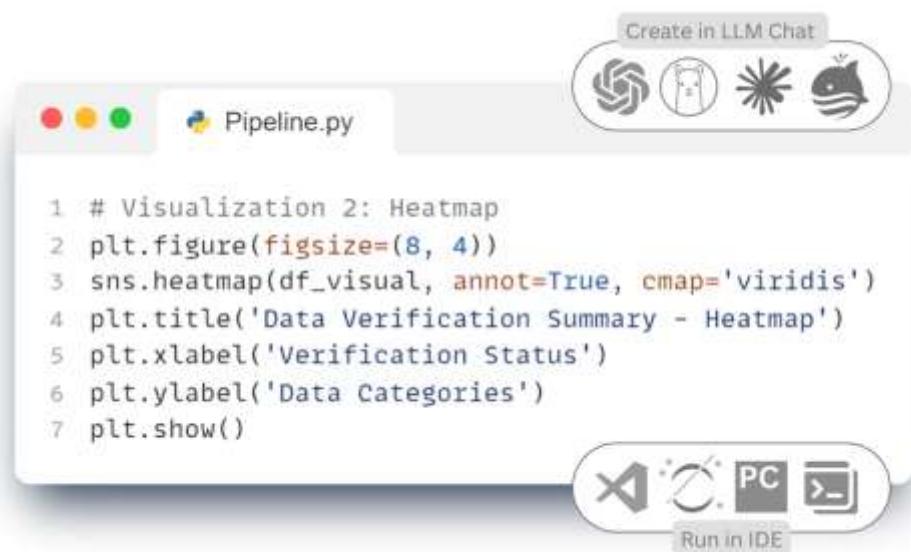
Znalost konkrétních knihoven pro vizualizaci není nutná - moderní nástroje, včetně LLM, umožňují na základě popisu úlohy automaticky generovat kód grafů i celých aplikací.

Výběr nástroje závisí na cílech projektu: zda se jedná o zprávu, prezentaci nebo online informační panel. Například open source knihovna Seaborn je obzvláště vhodná pro práci s kategorickými daty, pomáhá identifikovat vzory a trendy.

- Chcete-li se podívat na provoz knihovny Seaborn, můžete buď přímo požádat LLM o použití požadované knihovny, nebo zaslat podobný textový požadavek na adresu LLM:

Zobrazit tepelnou mapu ↗

- Odpověď LLM v podobě kódu a hotového grafu, jehož vykreslovací kód lze nyní zkopirovat do prostředí IDE a samotný graf lze zkopirovat nebo uložit pro vložení do dokumentu:



```

1 # Visualization 2: Heatmap
2 plt.figure(figsize=(8, 4))
3 sns.heatmap(df_visual, annot=True, cmap='viridis')
4 plt.title('Data Verification Summary - Heatmap')
5 plt.xlabel('Verification Status')
6 plt.ylabel('Data Categories')
7 plt.show()

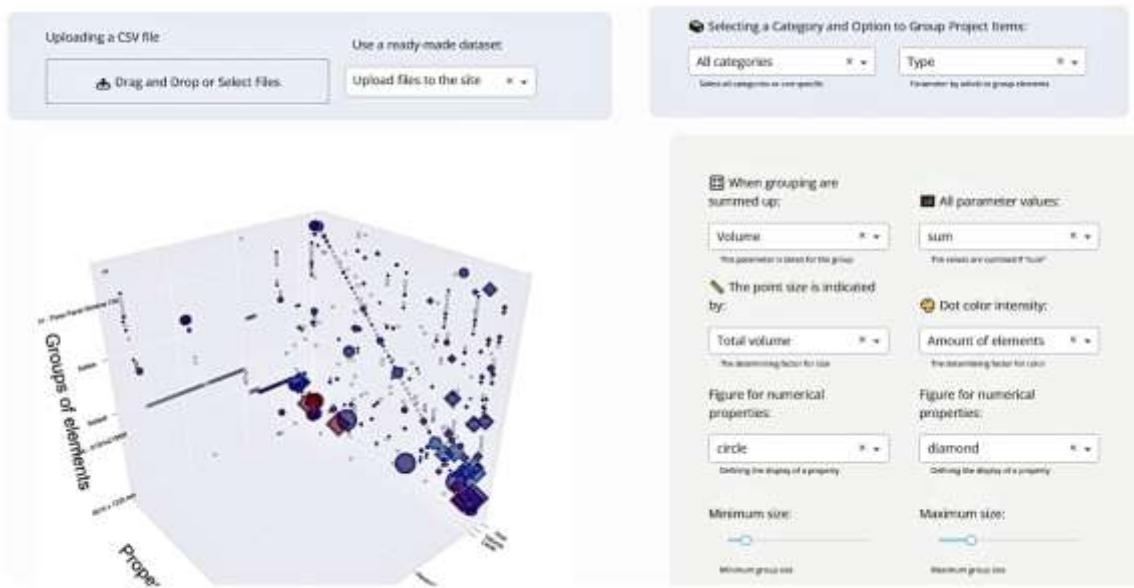
```

Run in IDE



Obrázek 7.2-11 Vizualizace výsledků validace (obrázek 7.2-8) dat pomocí knihovny Seaborn.

Pro ty, kteří dělají přednost interaktivnímu přístupu, existují nástroje, které umožňují vytvářet dynamické diagramy a panely s možností interakce. Knihovna Plotly (obrázek 7.1-6, obrázek 7.2-12) nabízí možnost vytvářet vysoce interaktivní grafy a panely, které lze vložit do webových stránek a umožnit uživateli interakci s daty v reálném čase.



Obr. 7.2-12 Interaktivní 3D -vizualizace atributů prvků z projektu CAD- (BIM-) pomocí knihovny Plotly.

Specializované knihovny s otevřeným zdrojovým kódem Bokeh, Dash a Streamlit poskytují pohodlný způsob prezentace dat, aniž by vyžadovaly hluboké znalosti vývoje webových stránek. Knihovna Bokeh je vhodná pro tvorbu složitých interaktivních grafů, Dash slouží k vytváření plnohodnotných analytických panelů a Streamlit umožňuje rychle vytvářet webové aplikace pro analýzu dat.

Díky těmto vizualizačním nástrojům mohou vývojáři a analytici efektivně šířit výsledky mezi kolegy a zúčastněné strany, což umožňuje intuitivní interakci s daty a zjednoduší rozhodování.

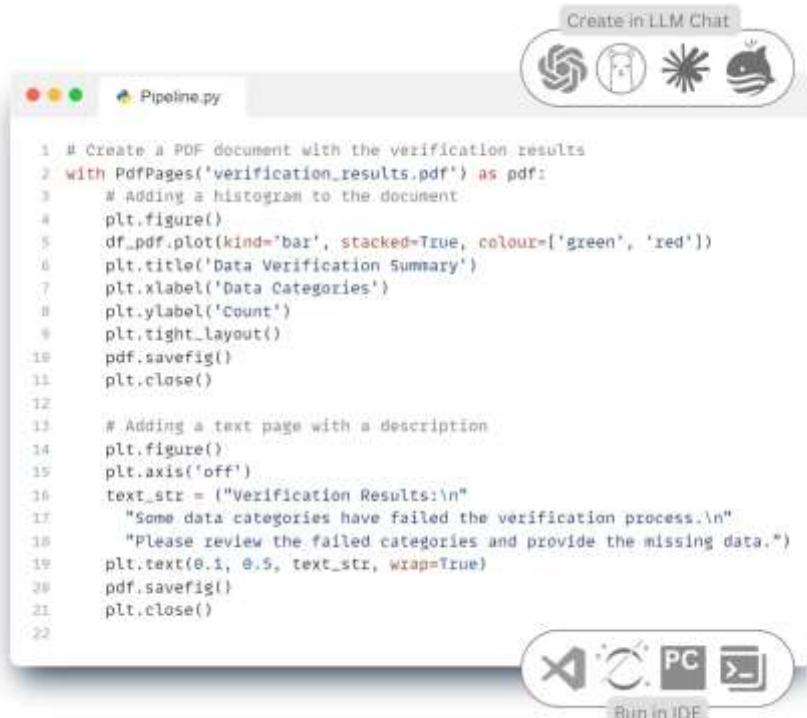
ETL Load: Automatické vytváření dokumentů PDF.

Ve fázi načítání dat můžete data nejen vizualizovat, nahrávat do tabulek nebo databází, ale také automaticky generovat reporty včetně potřebných grafů, tabulek a klíčových analytických ukazatelů, které obdrží manažer nebo specialista čekající na výsledky auditu. Automatické reporty mohou obsahovat jak komentáře, tak textovou interpretaci dat a vizualizační prvky - tabulky, grafy.

- 💡 Pro vytvoření PDF -zprávy s histogramem (obr. 7.2-9) a popisem analýzy na základě validace, kterou jsme provedli v předchozích kapitolách, stačí formulovat požadavek v pokračování dialogu s LLM, např.:

Napište kód pro vytvoření souboru PDF -file s histogramem a popisem výsledků výše uvedené validace dat (v chatu) a napište textové upozornění, že některé kategorie nebyly validovány a že je třeba doplnit chybějící. ↴

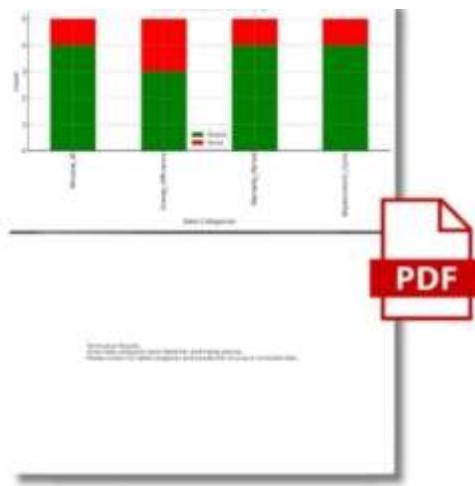
- 💡 Odpověď LLM jako kód a hotové PDF s výsledky:



```

1 # Create a PDF document with the verification results
2 with PdfPages('verification_results.pdf') as pdf:
3     # Adding a histogram to the document
4     plt.figure()
5     df.pdf.plot(kind='bar', stacked=True, colour=['green', 'red'])
6     plt.title('Data Verification Summary')
7     plt.xlabel('Data Categories')
8     plt.ylabel('Count')
9     plt.tight_layout()
10    pdf.savefig()
11    plt.close()
12
13    # Adding a text page with a description
14    plt.figure()
15    plt.axis('off')
16    text_str = ("Verification Results:\n"
17                "Some data categories have failed the verification process.\n"
18                "Please review the failed categories and provide the missing data.")
19    plt.text(0.1, 0.5, text_str, wrap=True)
20    pdf.savefig()
21    plt.close()
22

```



Obrázek 7.2-13 Automatický kód vytvoří PDF -dokument obsahující histogram s daty testu a text s výsledky testu.

Automaticky napsané řešení o pouhých 20 řádcích kódu pomocí LLM okamžitě vytvoří požadovaný dokument PDF (nebo DOC) vizualizací v podobě atributového histogramu (obr. 7.2-13) zobrazujícího počet dat, která prošla a neprošla validací, a s doplněním textového bloku shrnujícího výsledky a doporučení další postup.

Automatizované generování dokumentů je klíčovým prvkem fáze Load, zejména v projektovém prostředí, kde jsou rychlosť vykazování a přesnost kritické.

ETL Load: automatické generování dokumentů z FPDF.

Automatizace reportování ve fázi ETL Load je důležitým krokem při zpracování dat, zejména pokud je třeba výsledky analýz prezentovat ve formátu, který je snadno sdělitelný a srozumitelný. Ve stavebnictví je to často relevantní pro zprávy o postupu prací, statistiky projektových dat, zprávy o zajištění kvality nebo finanční dokumentaci.

Jedním z nejvhodnějších nástrojů pro takové úlohy je open source knihovna, FPDF, dostupná pro Python i PHP.

Knihovna **FPDF** s otevřeným zdrojovým kódem poskytuje flexibilní způsob generování dokumentů prostřednictvím kódu, který umožňuje přidávat záhlaví, text, tabulky a obrázky. Použití kódu namísto ručních úprav snižuje počet chyb a urychluje proces přípravy zpráv ve formátu PDF.

Jedním z klíčových kroků při vytváření dokumentu PDF je přidání nadpisů a hlavního textu ve formě komentářů nebo popisů. Při vytváření zprávy je však důležité nejen přidat text, ale také jej správně strukturovat. Nadpisy, odrážky, rádkování - to vše ovlivňuje čitelnost dokumentu. Pomocí FPDF, můžete nastavit parametry formátování, kontrolovat uspořádání prvků a přizpůsobit styl dokumentu.

FPDF je v principu velmi podobný HTML. Ti, kteří již znají HTML, mohou pomocí FPDF snadno vytvářet

dokumenty PDF jakékoli složitosti, protože struktura kódu je velmi podobná značkování v HTML: záhlaví, text, obrázky a tabulky se přidávají podobným způsobem. Ti, kteří HTML neznají, se nemusí obávat - mohou použít LLM, který vám okamžitě pomůže sestavit kód pro vygenerování požadovaného rozvržení dokumentu.

- Následující příklad ukazuje, jak vygenerovat sestavu se záhlavím a textem těla. Spuštěním tohoto kódu v libovolném vývojovém prostředí s podporou jazyka Python se vytvoří soubor PDF -file obsahující požadované záhlaví a text:

```
from fpdf import FPDF      # Import knihovny FPDF
pdf = FPDF()    # Vytvořte PDF -dokument
pdf.add_page()   # Přidání stránky

pdf.set_font("Arial", style='B', size=16) # Nastavení písma: Arial, tučné, velikost 16
pdf.cell(200, 10, "Zpráva o projektu", ln=True, align='C') # Vytvořte nadpis a zarovnejte jej na střed
pdf.set_font("Arial", size=12) # Změňte písmo na běžný Arial, velikost 12
pdf.multi_cell(0, 10, "Tento dokument obsahuje údaje o výsledcích ověření souboru projektu . . .") # Přidejte víceřádkový text
pdf.output(r "C:\reports\report.pdf") # Uložit PDF -file
```



Obrázek 7.2-14 Pomocí několika řádků kódu Pythonu můžeme automaticky vygenerovat textový dokument PDF, který potřebujeme.

Při přípravě zpráv je důležité vzít v úvahu, že údaje, z nichž je dokument vytvořen, jsou zřídkakdy statické. Záhlaví, textové bloky (obr. 7.2-14) jsou často tvoreny dynamicky, dostávají hodnoty ve fázi Transformace v procesu ETL.

Použití kódu umožňuje vytvářet dokumenty, které obsahují aktuální informace: název projektu, datum vytvoření zprávy, jakož i informace o účastnících nebo aktuálním stavu. Použití proměnných v kódu umožňuje automaticky vkládat tyto údaje na požadovaná místa ve zprávě, čímž zcela odpadá nutnost ruční úpravy před odesláním.

Vedle jednoduchého textu a nadpisů zaujímají v projektové dokumentaci zvláštní místo tabulky. Téměř každý dokument obsahuje strukturovaná data: od popisů objektů až po výsledky kontrol. Automatické generování tabulek na základě dat z fáze Transformace umožňuje nejen urychlit proces přípravy dokumentu, ale také minimalizovat chyby při přenosu informací. FPDF umožňuje vkládat tabulky do

PDF -souborů (jako text nebo obrázky), nastavovat okraje buněk, velikosti sloupců a písma (obr. 7.2-15). To je výhodné zejména při práci s dynamickými daty, kdy se počet řádků a sloupců může měnit v závislosti na úlohách dokumentu.

- Následující příklad ukazuje, jak automatizovat vytváření tabulek, např. se seznamy materiálů, odhady nebo výsledky testů parametrů:

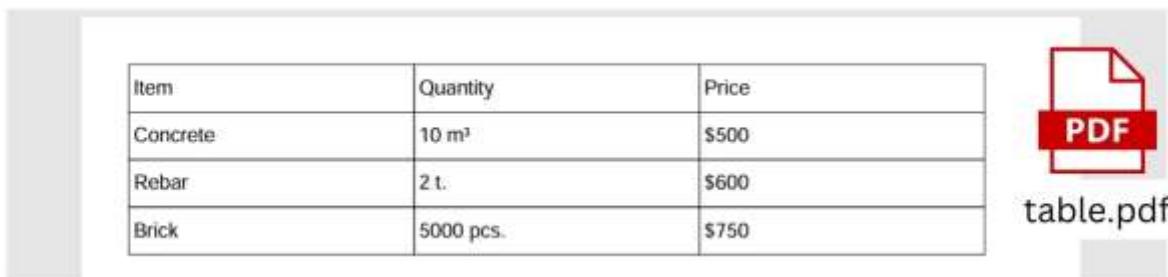
```

data = [
    ["Item", "Quantity", "Price"], # Nadpisy sloupců
    ["Beton", "10 m3", "$ 500."], # Údaje prvního řádku
    ["Rebar", "2 tuny", "$ 600"], # Údaje druhého řádku.
    ["Cihla", "5 000 kusů", "$ 750."], # Údaje třetího řádku.
]

pdf = FPDF () # Vytvoříme PDF -dokument
pdf.add_page() # Přidání stránky
pdf.set_font("Arial", size=12) # Nastavení písma

for row in data: # Prohledat řádky tabulky
    for item in row: # Projděte buňky v řádku
        pdf.cell(60, 10, item, border=1) # Vytvořte buňku s okrajem, šířkou 60 a výškou 10
    pdf.ln() # Přesun na další řádek
pdf.output(r "C:\reports\table.pdf") # Uložit PDF -file

```



Item	Quantity	Price
Concrete	10 m ³	\$500
Rebar	2 t.	\$600
Brick	5000 pcs.	\$750

Obrázek 7.2-15 Z kroku Transformace můžete automaticky generovat nejen text, ale také libovolné informace o tabulce v PDF.

V reálných scénářích vykazování jsou tabulky obvykle dynamicky generované informace získané ve fázi transformace dat. V uvedeném příkladu (obr. 7.2-15) je tabulka vložena do PDF -dokumentu ve statické podobě: data pro příklad byla umístěna do datového slovníku (první řádek kódu), v reálných podmírkách se taková datová proměnná doplní automaticky např. po seskupení datového rámce.

V praxi se takové tabulky často vytvářejí na základě strukturovaných dat pocházejících z různých dynamických zdrojů: databází, souborů aplikace Excel, rozhraní API nebo výsledků analytických výpočtů. Nejčastěji se ve fázi transformace (ETL) data agregují, seskupují nebo filtroují - a teprve poté se transformují do souhrnů v podobě grafů nebo dvourozměrných tabulek zobrazovaných v sestavách. To znamená, že obsah tabulky se může měnit v závislosti na zvolených parametrech, období analýzy, projektových filtroch nebo uživatelských nastaveních.

Díky použití dynamických datových rámců a datových sad ve fázi Transform je proces vykazování ve fázi Load maximálně flexibilní, škálovatelný a snadno opakovatelný bez nutnosti manuálního zásahu.

Kromě tabulek a textu podporuje FPDF také přidávání grafů tabulkových dat, což umožňuje vkládat do sestavy obrázky vygenerované pomocí Matplotlib nebo jiných vizualizačních knihoven, o kterých jsme hovořili výše. Pomocí kódu lze do dokumentu přidat libovolné grafy, diagramy a schémata.

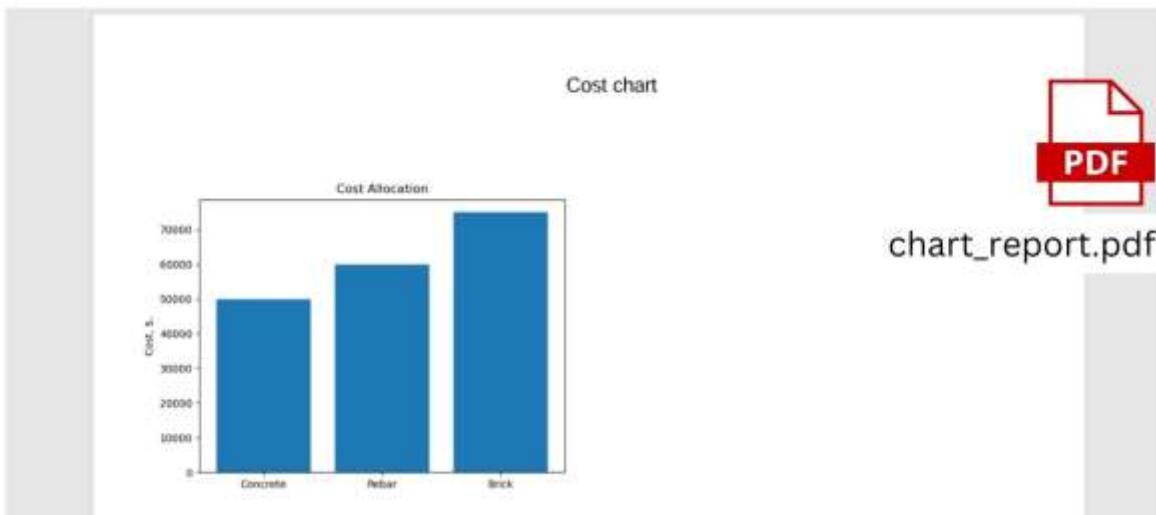
- Pomocí knihovny Pythonu FPDF, přidáme do dokumentu PDF graf předgenerovaný pomocí Matplotlib:

```
import matplotlib.pyplot as plt # Import Matplotlib pro vytváření grafů

fig, ax = plt.subplots() # Vytvoření obrázku a os grafu
categories = ["Concrete", "Rebar", "Brick"] # Názvy kategorií
values = [50000, 60000, 75000] # Hodnoty kategorií
ax.bar(categories, values) # Vytvoří sloupcový graf
plt.ylabel("Hodnota,$.") # Označte osu Y
plt.title("Rozdělení nákladů") # Přidejte nadpis
plt.savefig(r "C:\reports\chart\chart\chart.png") # Uložte graf jako obrázek

pdf = FPDF () # Vytvoří PDF -dokument
pdf.add_page() # Přidat stránku
pdf.set_font("Arial", size=12) # Nastavte písmo
pdf.cell(200, 10, "Cost Chart", ln=True, align='C') # Přidejte záhlaví

pdf.image(r "C:\reports\chart\chart\chart.png", x=10, y=30, w=100) # Vložte obrázek do PDF (x, y - souřadnice, w - šířka)
pdf.output(r "C:\reports\chart_report.pdf") # Uložte soubor PDF
```



Obr. 7.2-16 Pomocí několika řádků kódu můžete vygenerovat graf, uložit jej a poté vložit do dokumentu PDF.

S FPDF se proces přípravy dokumentů a logiky stává transparentním, rychlým a pohodlným. Šablony zabudované v kódu umožňují generovat dokumenty s aktuálními údaji, čímž odpadá nutnost ručního vyplňování.

Použití automatizace ETL - místo časově náročného manuálního reportování se mohou profesionálové soustředit na analýzu dat a rozhodování, nikoli na výběr správného nástroje pro práci s konkrétním datovým silem s přehledným uživatelským rozhraním.

Knihovna FPDF tak poskytuje flexibilní nástroj pro automatizovanou tvorbu dokumentů libovolné složitosti - od krátkých technických zpráv až po složité analytické souhrny s tabulkami a grafy, což umožňuje nejen urychlit tok dokumentů, ale také výrazně snížit pravděpodobnost chyb spojených s ručním zadáváním a formátováním dat.

ETL Load: Reportování a načítání do jiných systémů.

Ve fázi Load byly výsledky generovány ve formě tabulek, grafů a závěrečných zpráv ve formátu PDF, které byly připraveny v souladu se stanovenými požadavky. Dále je možné tato data exportovat do strojově čitelných formátů (např. CSV), což je nezbytné pro integraci s externími systémy, jako jsou ERP, CAFM, CPM, BI platformy a další podniková nebo oborová řešení. Kromě formátu CSV lze data nahrávat do formátů XLSX, JSON, XML nebo přímo do databází, které podporují automatickou výměnu informací.

- 💡 Chcete-li vygenerovat příslušný kód pro automatizaci kroku Load, jednoduše se zeptejte například na rozhraní LLM -interface: ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude nebo QWEN:

Napište kód pro generování sestavy výsledků validace dat v DataFrame, kde jsou sloupce s předponou "ověřeno_" spočítány, přejmenovány na "Vyhověl" a "Nevyhověl", chybějící hodnoty jsou nahrazeny hodnotou 0 a poté jsou do souboru CSV exportovány pouze ty řádky, které prošly všemi validacemi.

 Odpověď společnosti LLM:



```

1 # Create a report on the results of data validation
2 report = df.filter(like='verified_').apply(lambda x: x.value_counts());T
3 report.columns = ['Passed', 'Failed']
4 report.fillna(0, inplace=True)
5
6 # Export the cleaned and verified data to a CSV file
7 verified_df = df[df.filter(like='verified_').all(axis=1)]
8 verified_df.to_csv('verified_data.csv', index=False)

```

Obrázek 7.2-17 Ověřená data získaná v kroku Transformace z konečného datového rámce se exportují do souboru CSV - pro integraci s jinými systémy.

Výše uvedený kód (obr. 7.2-17) implementuje závěrečnou fázi ETL -procesu - Load, během níž jsou ověřená data uložena ve formátu CSV, kompatibilním s většinou externích systémů a databází. Tím jsme dokončili celý cyklus ETL procesu, včetně extrakce, transformace, vizualizace, dokumentace a exportu dat do potřebných systémů a formátů, což zajišťuje reprodukovatelnost, transparentnost a automatizaci práce s informacemi.

Potrubí ETL lze použít jak pro zpracování jednotlivých projektů, tak pro rozsáhlé aplikace - při analýze stovek nebo tisíců příchozích dat v podobě dokumentů, obrázků, skenů, projektů CAD, mračen bodů, souborů PDF nebo jiných zdrojů pocházejících z distribuovaných systémů. Díky možnosti plně automatizovat proces není ETL pouze technickým nástrojem pro zpracování, ale základem digitální informační infrastruktury ve stavebnictví.

ETL s LLM: Vizualizace dat z PDF -dokumentů.

Je čas přejít k vytvoření plnohodnotného procesu ETL, který zahrnuje všechny klíčové fáze zpracování dat v jediném scénáři - extrakci, transformaci a načítání. Vytvořme automatizovaný ETL-Pipeline, který vám umožní zpracovávat dokumenty PDF bez manuální práce - extrahovat data z dokumentů, vizualizovat je, analyzovat a přenášet do jiných systémů.

Proces ETL bude v našem příkladu popsán pomocí výzvy, které budou muset jazykovému modelu

(LLM) vysvětlit všechny procesy ETL s popisem konečného výsledku, který má být získán. V tomto případě je úkolem vyhledat všechny soubory PDF v zadané složce a jejích podsložkách, extrahovat z nich relevantní informace - např. názvy materiálů, množství a náklady - a výsledek prezentovat jako strukturovanou tabulkou (DataFrame) pro další analýzu

- 💡 První textový dotaz v LLM k automatickému extrahování dat z více dokumentů PDF a vytvoření datového rámce dat pro krok Extract:

Napište kód pro extrakci informací o materiálu ze souborů PDF v dané složce a jejích podsložkách. Údaje v PDF obsahují název materiálu, množství a cenu. Výsledek by měl být uložen v DataFrame.

- 💡 Odpověď společnosti LLM :



The screenshot shows a user interface for generating Python code. At the top, there's a button labeled "Create in LLM Chat". Below it is a toolbar with icons for a brain, a document, a star, and a whale. The main area is a code editor window titled "Pipeline.py" containing the following Python script:

```

1 import os
2 import pdfplumber
3 import pandas as pd
4
5 def extract_data(base_folder_path):
6     """
7         Extract material data from PDF files in the specified base
8         folder and all its subfolders. Assumes each line in the PDF contains
9         'Material' followed by material name, quantity, and cost.
10    """
11    data = []
12    # Walk through each folder and subfolder in the base folder
13    for folder_path, _, files in os.walk(base_folder_path):
14        for file in filter(lambda f: f.endswith('.pdf'), files):
15            with pdfplumber.open(os.path.join(folder_path, file)) as pdf:
16                # Process each page in the PDF
17                for page in pdf.pages:
18                    # Splitting text into lines
19                    for line in page.extract_text().split('\n'):
20                        # Check if the line contains material data
21                        if 'Material' in line:
22                            parts = line.split()
23                            # Append data to list; convert quantity and cost to float
24                            data.append({'file': file, 'material': parts[1],
25                                         'quantity': float(parts[2]), 'cost': float(parts[3])})
26
27    # Convert list of dictionaries to pandas DataFrame
28    return pd.DataFrame(data)
29
30 # Specify the base folder path containing PDF files and subfolders with PDFs
31 df = extract_data('C:\\')

```

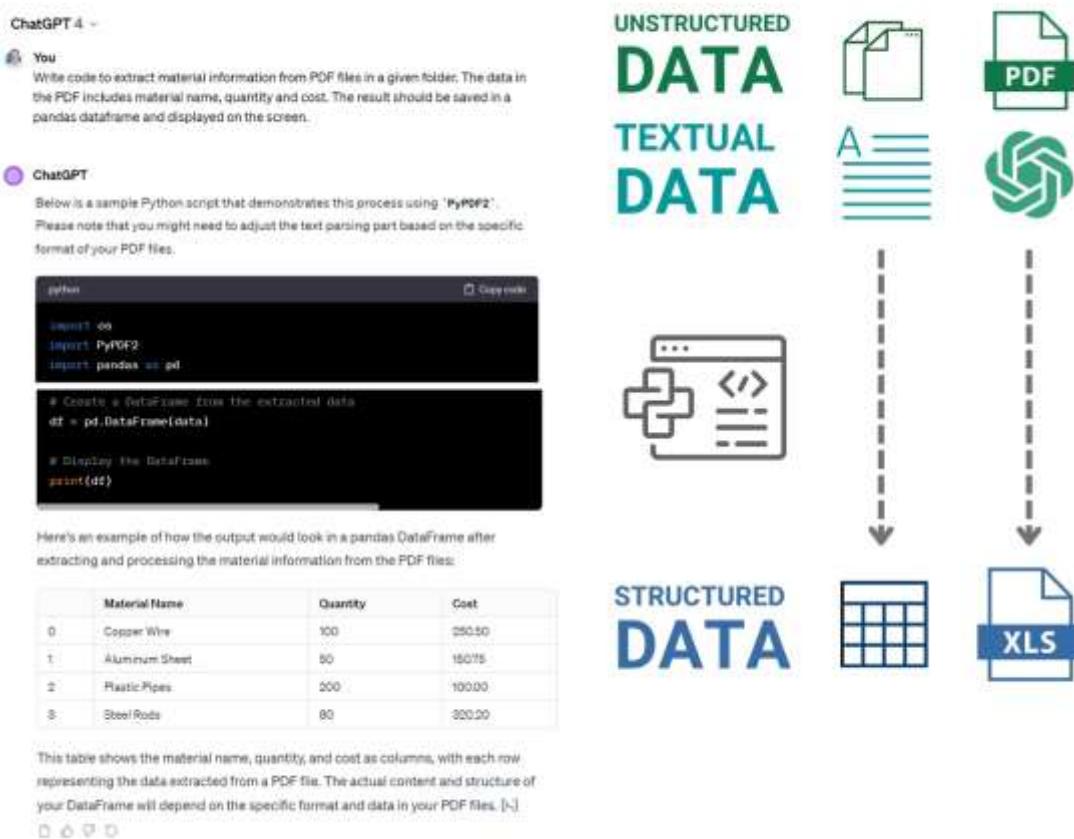
At the bottom of the code editor, there's a toolbar with icons for a file, a circular arrow, a PC, and a clipboard, and a button labeled "Run in IDE".

Obrázek 7.2-18 LLM vytvoří kód jazyka Python -kód pro extrakci dat ze souborů PDF - souborů v určité složce a všech jejích podsložkách.

Odpověď LLM (obrázek 7.2-18) je hotový skript v jazyce Python -skript, který automaticky prochází

všechny složky, otevírá nalezené soubory PDF -soubory, extrahuje z nich textové informace a převádí je do tabulky. Výsledný kód chatu lze spustit v některém z populárních IDE PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse s pluginem PyDev, Thonny, Wing IDE, IntelliJ IDEA s pluginem Python, JupyterLab nebo populárních online nástrojích Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Alternativně, pro zjednodušení procesu sběru, můžeme místo kopírování kódu z LLM a používání kódu v IDE také nahrát tucet PDF -souborů přímo do chatu LLM (obrázek 7.2-19) a získat jako výstup tabulku, aniž bychom museli kód vidět nebo spouštět. Výsledkem spuštění tohoto kódu bude tabulka s námi vybranými atributy.

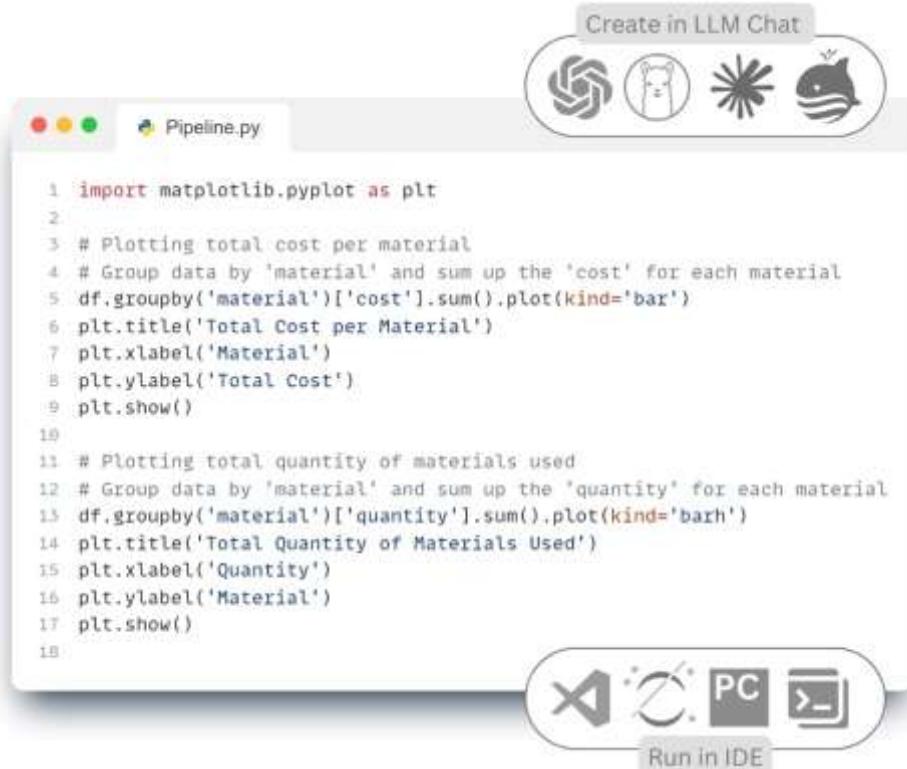


Obrázek 7.2-19 Výsledek spuštění kódu v LLM, který extrahuje data ze souborů PDF - v zobrazení strukturovaného datového rámce s vybranými atributy.

V dalším kroku požádáme o jazykový model na získaných datech - například pro porovnání nákladů a objemu spotřeby materiálu a vytvoření několika příkladů vizualizací, které budou sloužit jako podklad pro další analýzu.

- Požádejte v pokračujícím rozhovoru s LLM o vykreslení některých grafů z tabulek, které byly vytvořeny v kroku Transformace (obrázek 7.2-18):

Vizualizace celkových nákladů a množství každého materiálu z DataFrame (obrázek 7.2-18) ↴



The screenshot shows a LLM interface with a "Create in LLM Chat" button and four icons (chatbot, cat, asterisk, whale). Below is a code editor window titled "Pipeline.py" containing the following Python code:

```

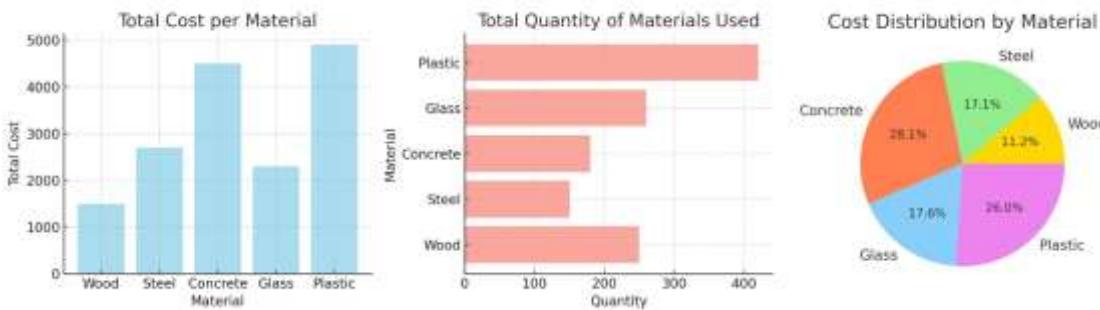
1 import matplotlib.pyplot as plt
2
3 # Plotting total cost per material
4 # Group data by 'material' and sum up the 'cost' for each material
5 df.groupby('material')['cost'].sum().plot(kind='bar')
6 plt.title('Total Cost per Material')
7 plt.xlabel('Material')
8 plt.ylabel('Total Cost')
9 plt.show()
10
11 # Plotting total quantity of materials used
12 # Group data by 'material' and sum up the 'quantity' for each material
13 df.groupby('material')['quantity'].sum().plot(kind='barh')
14 plt.title('Total Quantity of Materials Used')
15 plt.xlabel('Quantity')
16 plt.ylabel('Material')
17 plt.show()
18

```

At the bottom right is a "Run in IDE" button with icons for VS Code, Python, and terminal.

Obrázek 7.2-20 Odezva LLM -modely jako kód Pythonu pro vizualizaci dat z datového rámce pomocí knihovny matplotlib.

LLM automaticky generuje a spouští Python -kód (obr. 7.2-20) pomocí knihovny matplotlib. Po spuštění tohoto kódu získáme grafy nákladů a spotřeby materiálu ve stavebních projektech přímo v chatu (obr. 7.2-21), což značně zjednoduší analytickou práci.

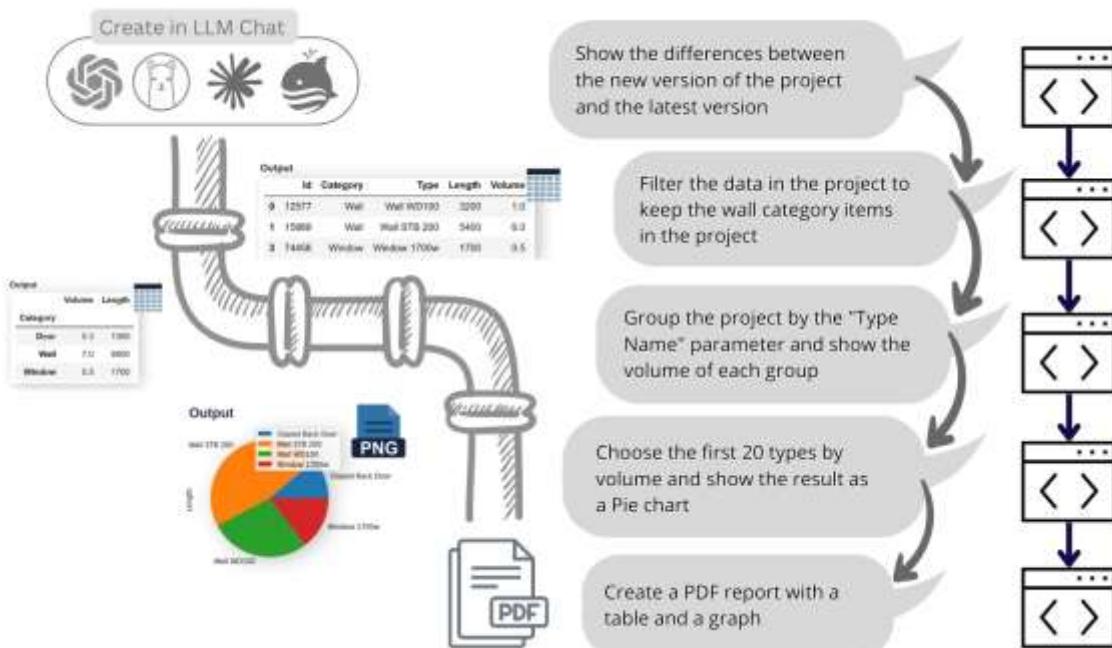


Obr. 7.2-21 Vizualizace odezvy LLM v podobě grafů na základě dat shromážděných v DataFrame.

Podpora při vytváření nápadů na psaní kódu ETL, analýze a provádění kódu a vizualizaci výsledků je k

dispozici prostřednictvím jednoduchých textových dotazů v LLM, aniž by bylo nutné se učit základy programování. Vznik nástrojů umělé inteligence, jako je LLM, rozhodně mění přístup k programování a automatizaci zpracování dat (obrázek 7.2-22).

Podle zprávy PwC "Jaká je skutečná hodnota umělé inteligence pro váš podnik a jak ji můžete využít?". (2017) [139] budou hlavními tahouny hospodářského růstu automatizace procesů a zvyšování produktivity. A očekává se, že zlepšení produktivity bude v letech 2017 až 2030 představovat více než 55% veškerého růstu HDP způsobeného umělou inteligencí."



Obrázek 7.2-22 AI LLM pomáhá generovat návrh kódu, který se použije v budoucích projektech bez nutnosti LLM.

Pomocí nástrojů, jako jsou ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok, jakož i otevřených dat a softwaru s otevřeným zdrojovým kódem, můžeme automatizovat procesy, které se dříve prováděly pouze pomocí specializovaných, nákladných a obtížně udržovatelných modulárních proprietárních systémů.

V kontextu stavebnictví to znamená, že společnosti, které jako první zavedou automatizované procesy Pipeline -data, získají významné výhody, od vyšší efektivity řízení projektů přes snížení finančních ztrát až po odstranění roztríštěných aplikací a oddělených datových skladů.

Popsaná logika provádění obchodních úloh v procesu ETL je klíčovou součástí automatizace analytických a procesů zpracování dat, která je specifickou variantou širšího pojmu - potrubí (Pipelines).

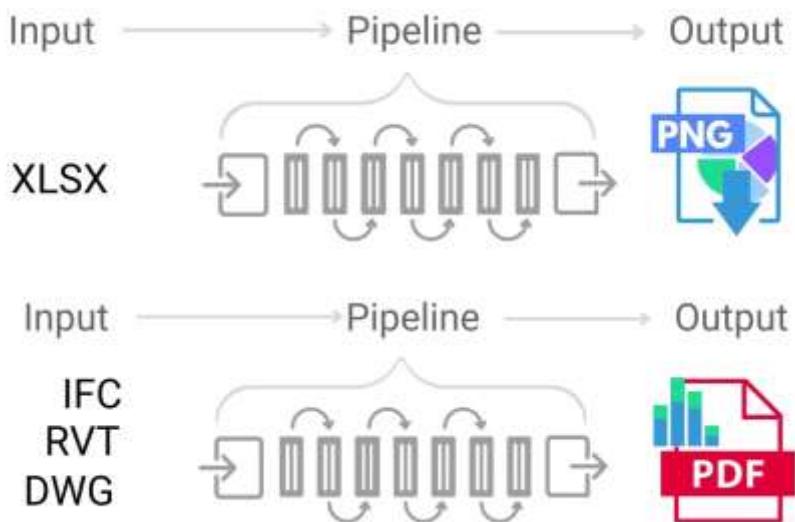
KAPITOLA 7.3.

AUTOMATICKÝ ETL DOPRAVNÍK (POTRUBÍ

Potrubí: Automatický dopravník ETL data.

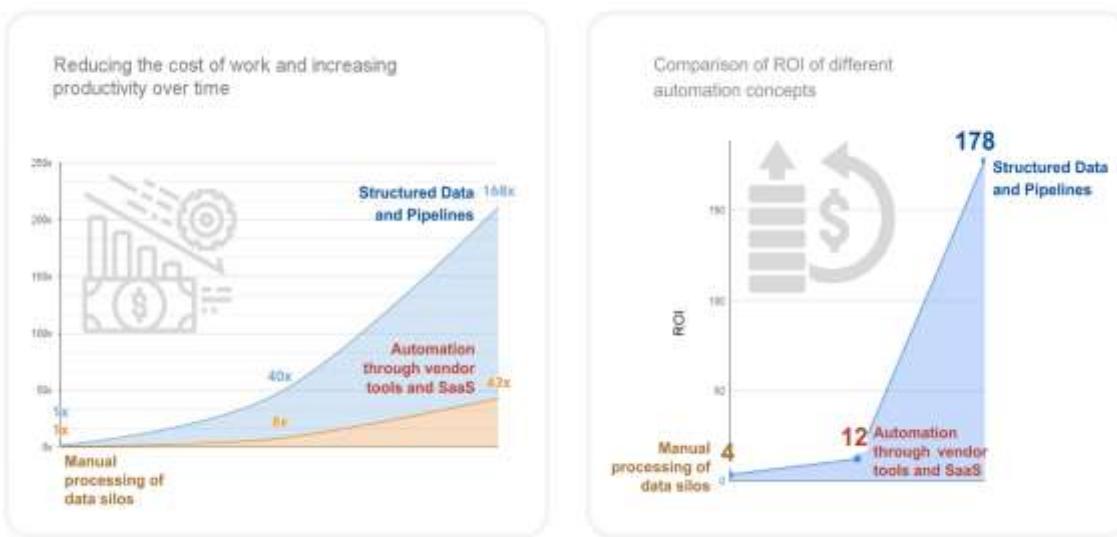
Proces ETL se tradičně používá ke zpracování dat v analytických systémech, které zahrnují strukturované i nestrukturované zdroje. V dnešním digitálním prostředí se však stále častěji používá širší termín - Pipeline (dopravník), který popisuje jakýkoli sekvenční řetězec zpracování, kde se výstup jedné fáze stává vstupem pro další.

Tento přístup se vztahuje nejen na data, ale i na další typy automatizace: zpracování úloh, vytváření zpráv, integraci se softwarem a digitální pracovní postupy (obr. 7.3-1).



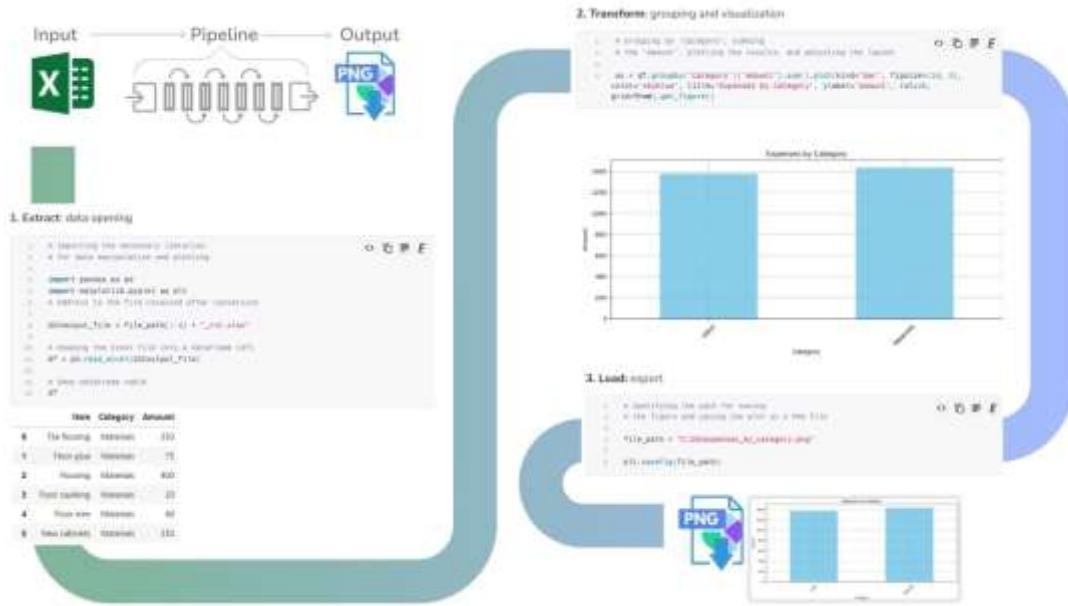
Obrázek 7.3-1 Pipeline je zpracovatelská sekvence, v níž se výstup jednoho stupně stává vstupem do dalšího stupně.

Použití pipeline je jedním z hlavních prvků automatizace, zejména při práci s velkým množstvím heterogenních dat. Architektura Pipeline umožňuje uspořádat složité kroky zpracování do modulárního, konzistentního a spravovatelného formátu, což zvyšuje čitelnost, zjednoduší údržbu kódu a umožňuje inkrementální ladění a škálovatelné testování.



Obr. 7.3-2 ROI Proces validace dat pomocí potrubí zkracuje dobu provádění o desítky a stovky hodin ve srovnání se zpracováním pomocí klasických nástrojů [74].

Na rozdíl od manuální práce v proprietárních systémech (ERP, PMIS, CAD, atd.) umožnuje pipelining výrazně (obr. 7.3-2) zvýšit rychlosť úloh, zamezit opakoványm činnostem a automatizovat spouštění procesů ve správný čas (obr. 7.3-3).

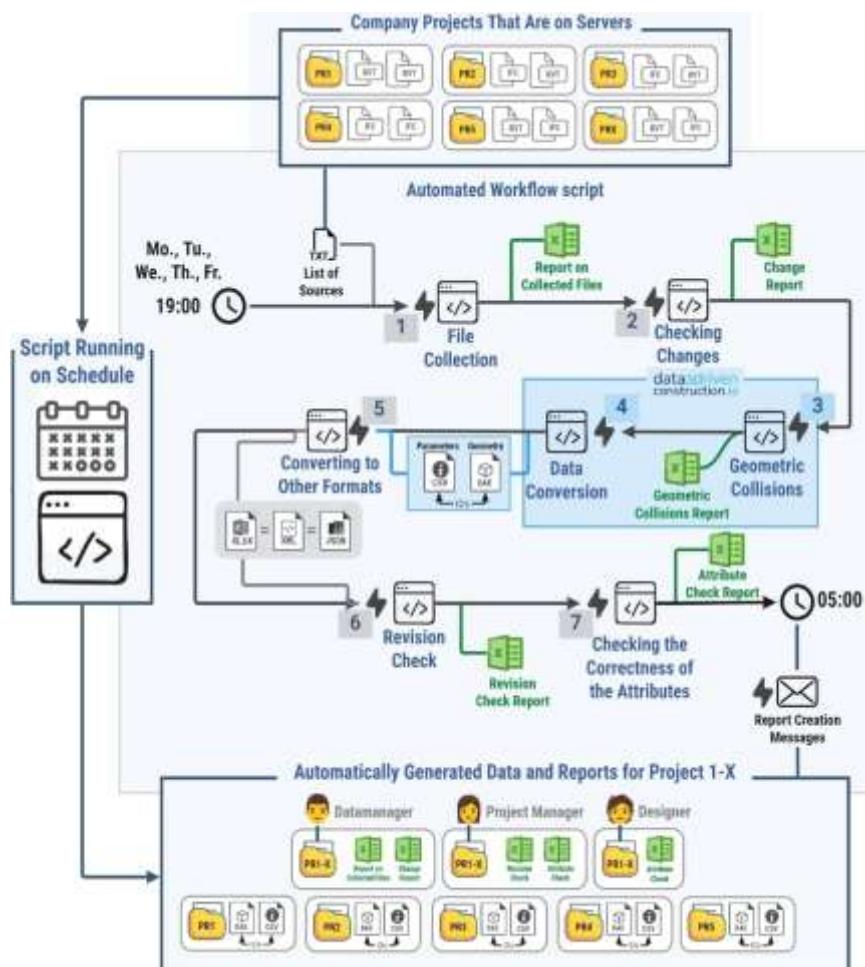


Obr. 7.3-3 Příklad ETL Potrubí na automatické načtení grafu z tabulkových dat v souboru XLSX bez otevření aplikace Excel.

Chcete-li zpracovávat proudová data a vytvořit automatizovaný Pipeline, podobně jako u procesu ETL, musíte předem určit zdroje dat a časový rámec jejich sběru - buď pro konkrétní obchodní proces, nebo pro celou společnost.

Ve stavebních projektech pocházejí data z mnoha různorodých zdrojů s různými intervaly aktualizace. Pro vytvoření spolehlivé datové vitriny je zásadní zaznamenávat, kdy jsou informace získávány a aktualizovány. To umožňuje včasné rozhodování a zvyšuje efektivitu řízení projektu.

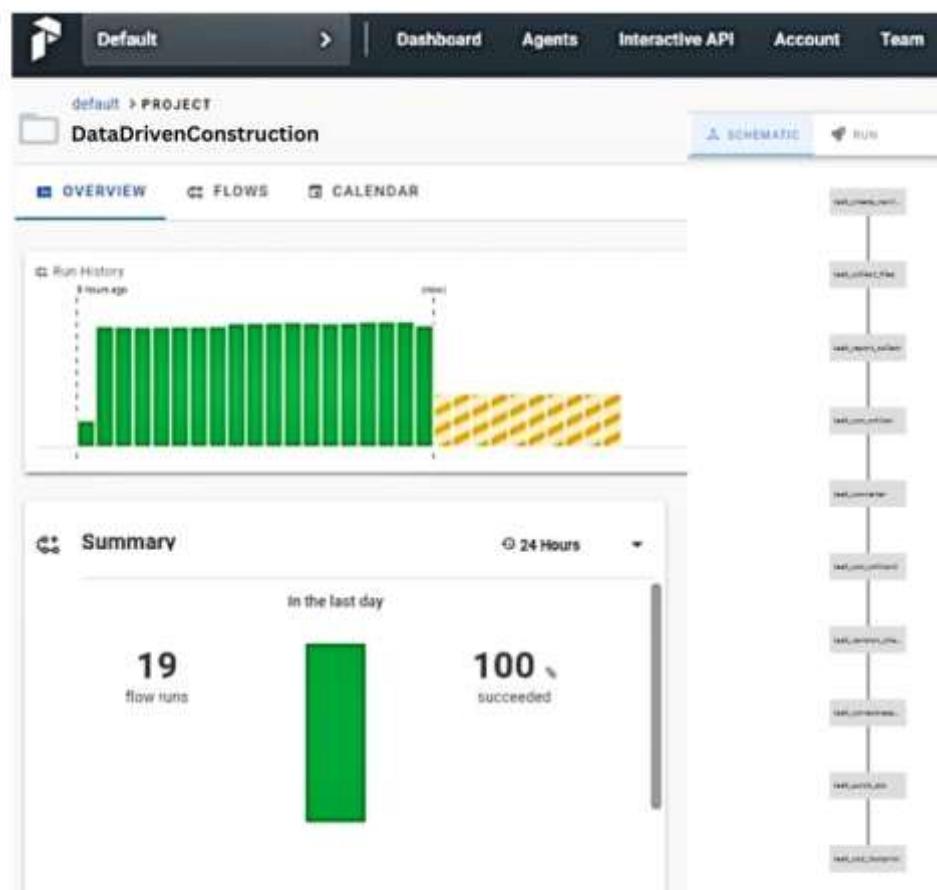
Jednou z možností je zahájit proces montáže v pevně stanovený čas - například v 19:00 na konci pracovního dne. V tomto okamžiku se aktivuje první skript odpovědný za agregaci dat z různých systémů a úložišť (obr. 7.3-4 krok 1). Následuje automatické zpracování a transformace dat do strukturovaného formátu vhodného pro analýzu (obr. 7.3-4 krok 2-4). V závěrečné fázi jsou z připravených dat automaticky generovány reporty, dashboardy a další produkty popsané v předchozích kapitolách (obr. 7.3-4 krok 6-7). Výsledkem je, že v 5:00 ráno již mají manažeři k dispozici aktuální zprávy o stavu projektu v požadovaném formátu (obr. 7.3-5).



Obr. 7.3-4 Data v systému Pipeline, která se automaticky shromažďují večer, se zpracovávají přes noc, takže manažeři mají ráno k dispozici aktuální zprávy a čerstvé reporty.

Včasný sběr dat, definice klíčových ukazatelů výkonnosti, automatizace transformačních procesů a vizualizace prostřednictvím řídicích panelů jsou klíčovými prvky úspěšného rozhodování založeného na datech.

Takové automatizované procesy (obr. 7.3-4) lze provádět zcela autonomně: běží podle plánu, zpracovávají data bez obsluhy a mohou být nasazeny buď v cloudu, nebo na vlastním serveru společnosti (obr. 7.3-5). Díky tomu lze takovéto pipeline ETL integrovat do stávající IT infrastruktury, zachovat kontrolu nad daty a zajistit flexibilitu při škálování.



Obr. 7.3-5 Automatické ETL -konvertor procesy (obr. 7.3-4) na platformě Prefect, ve které se každý pracovní den po 19:00 střídavě spouští 10 python skriptů.

Automatizace pracovních postupů nejenže zvyšuje produktivitu týmu tím, že uvolňuje čas na smysluplnější a méně rutinní úkoly, ale slouží také jako důležitý první krok k začlenění technologií umělé inteligence (AI) do podnikových procesů, o nichž se podrobněji zmíníme v kapitole o prediktivní analýze a strojovém učení.

Potrubí -ETL proces validace dat pomocí LLM

V předchozích kapitolách o vytváření požadavků na data a automatizaci ETL, jsme krok za krokem rozebrali proces přípravy, transformace, validace a vizualizace dat. Tyto činnosti byly implementovány jako samostatné bloky kódu (obrázek 7.2-18 - obrázek 7.2-20), z nichž každý vykonává určitou úlohu.

Nyní máme další cíl - spojit tyto prvky do jediné, ucelené a automatizované pipeline zpracování dat - pipeline, ETL -Pipeline - ve které se všechny fáze (načítání, validace, vizualizace, export) provádějí postupně v jediném automaticky spustitelném skriptu.

V následujícím příkladu bude realizován kompletní cyklus zpracování dat: od načtení zdrojového souboru CSV →po kontrolu struktury a hodnot pomocí regulárních výrazů →výpočet výsledků→ vygenerování vizuální zprávy ve formátu PDF.

💡 Pro získání příslušného kódu můžete použít následující textový dotaz do LLM:

Napište prosím ukázku kódu, který načte data z CSV, validuje DataFrame pomocí regulárních výrazů, kontroluje identifikátory ve formátu "W-NEW" nebo "W-OLD", energetickou účinnost s písmeny "A" až "G", záruční dobu a cyklus výměny s číselnými hodnotami v letech a na konci vytvoří sestavu s počtem vyhovujících a nevyhovujících hodnot, vygeneruje PDF s histogramem výsledků a přidá textový popis. ↵

■ Odpověď společnosti LLM :



```

1 # Importing necessary libraries
2 import pandas as pd
3 import re
4 import matplotlib.pyplot as plt
5 from matplotlib.backends.backend_pdf import PdfPages
6
7
8 # Load data
9 df = pd.read_csv('path_to_csv_file.csv')
10
11 # Define patterns and verify data
12 patterns = {
13     'Window_ID': r'^W-NEW|W-OLD[0-9]+|D-1[0-9]{2}\$',
14     'Energy_Efficiency': r'^[A-G]\+\$',
15     'Warranty_Period': r'^\d+\syears\$',
16     'Replacement_Cycle': r'^\d+\syears\$'
17 }
18
19 for column, pattern in patterns.items():
20     df['verified_' + column] = df[column].apply(lambda x: bool(re.match(pattern, str(x))))
21
22 # Create report
23 report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
24 report.columns = ['Passed', 'Failed']
25
26 # Generate PDF
27 pdf_path = '/mnt/data/verification_results_full_pipeline.pdf'
28 with PdfPages(pdf_path) as pdf:
29     # Plot and save bar chart
30     report.plot(kind='bar', stacked=True, color=['green', 'red'])
31     plt.title('Data Verification Summary')
32     plt.xlabel('Data Categories')
33     plt.ylabel('Count')
34     plt.tight_layout()
35     pdf.savefig()
36     plt.close()
37
38 # Add text page
39 plt.figure()
40 plt.axis('off')
41 text_str = "Verification Results Summary"
42 plt.text(0.1, 0.5, text_str, wrap=True)
43 pdf.savefig()
44 plt.close()
45
46 pdf_path
47

```

Obr. 7.3-6 Pipeline (ETL) automatizuje celý cyklus zpracování dat: od stažení a validace až po vytvoření strukturované zprávy ve formátu PDF .

Automatický kód (obr. 7.3-6) uvnitř chatu LLM nebo v DIE po zkopírování kódu ověří data ze souboru CSV -file pomocí zadaných regulárních výrazů, vytvoří zprávu o počtu vyhovujících a nevyhovujících záznamů a poté uloží výsledky ověření jako soubor PDF -file.

Tato struktura ETL -conveyor, kde je každý krok - od načítání dat až po generování sestav - implementován jako samostatný modul, zajišťuje transparentnost, škálovatelnost a reproducibilnost. Prezentace validační logiky jako snadno čitelného kódu v jazyce Python činí proces transparentním a srozumitelným nejen pro vývojáře, ale také pro specialisty na správu dat, kvalitu a analytiku.

Přístup společnosti Pipeline k automatizaci zpracování dat umožňuje standardizovat procesy, zvýšit jejich opakovatelnost a zjednodušit přizpůsobení novým projektům. Vzniká tak jednotná metodika analýzy dat bez ohledu na zdroj nebo typ úlohy - ať už jde o testování shody, reporting nebo přenos dat do externích systémů.

Taková automatizace snižuje počet lidských chyb, omezuje závislost na proprietárních řešeních a zvyšuje přesnost a spolehlivost výsledků, takže je vhodná jak pro operativní analýzu na úrovni projektů, tak pro strategickou analýzu na úrovni společnosti.

Potrubí -ETL: ověření dat a informací o prvcích projektu v CAD (BIM).

Data ze systémů CAD a databází (BIM) patří k nejpracovanějším a nejdynamičtěji aktualizovaným zdrojům dat ve stavebních firmách. Tyto aplikace nejenž popisují projekt pomocí geometrie, ale doplňují jej také o několik vrstev textových informací: objemy, vlastnosti materiálů, přiřazení místností, úrovně energetické účinnosti, tolerance, předpokládanou životnost a další atributy.

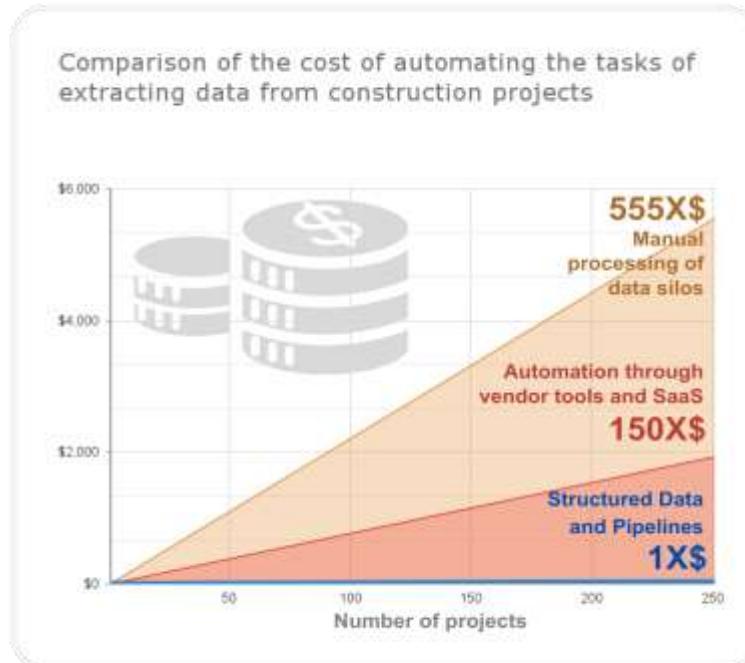
Atributy přiřazené entitám v CAD -modelu se vytvářejí ve fázi návrhu a stávají se základem pro další obchodní procesy, včetně kalkulace nákladů, plánování, hodnocení životního cyklu a integrace se systémy ERP a CAFM -systémy, kde efektivita procesů do značné míry závisí na kvalitě dat přicházejících z konstrukčních oddělení.

Tradiční přístup k validaci atributů v modelech CAD- (BIM-) zahrnuje ruční validaci (obr. 7.2-1), která se při velkém objemu modelů stává zdlouhavým a nákladným procesem. Vzhledem k objemu a počtu moderních stavebních projektů a jejich pravidelným aktualizacím se proces validace a transformace dat stává neudržitelným a cenově nedostupným.

Generální dodavatelé a projektoví manažeři se potýkají s nutností zpracovávat velké množství projektových dat, včetně více verzí a fragmentů stejných modelů. Data přicházejí od projekčních organizací ve formátech RVT, DWG, DGN, IFC, NWD a dalších (obrázek 3.1-14) a vyžadují pravidelnou kontrolu shody s oborovými a podnikovými normami

Závislost na manuálních činnostech a specializovaném softwaru činí z procesu validace dat úzké

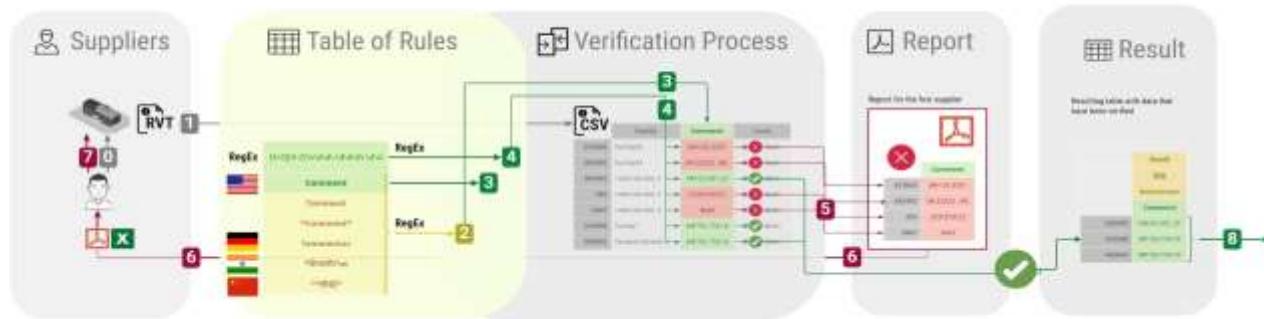
místo v pracovních postupech souvisejících s daty z celopodnikových modelů. Automatizace a používání strukturovaných požadavků může tuto závislost odstranit, čímž se výrazně zvýší rychlosť a spolehlivosť validace dat (obrázek 7.3-7).



Obr. 7.3-7 Automatizace zvyšuje rychlosť ověřování a zpracování dat, což snižuje náklady na práci desítkykrát [140].

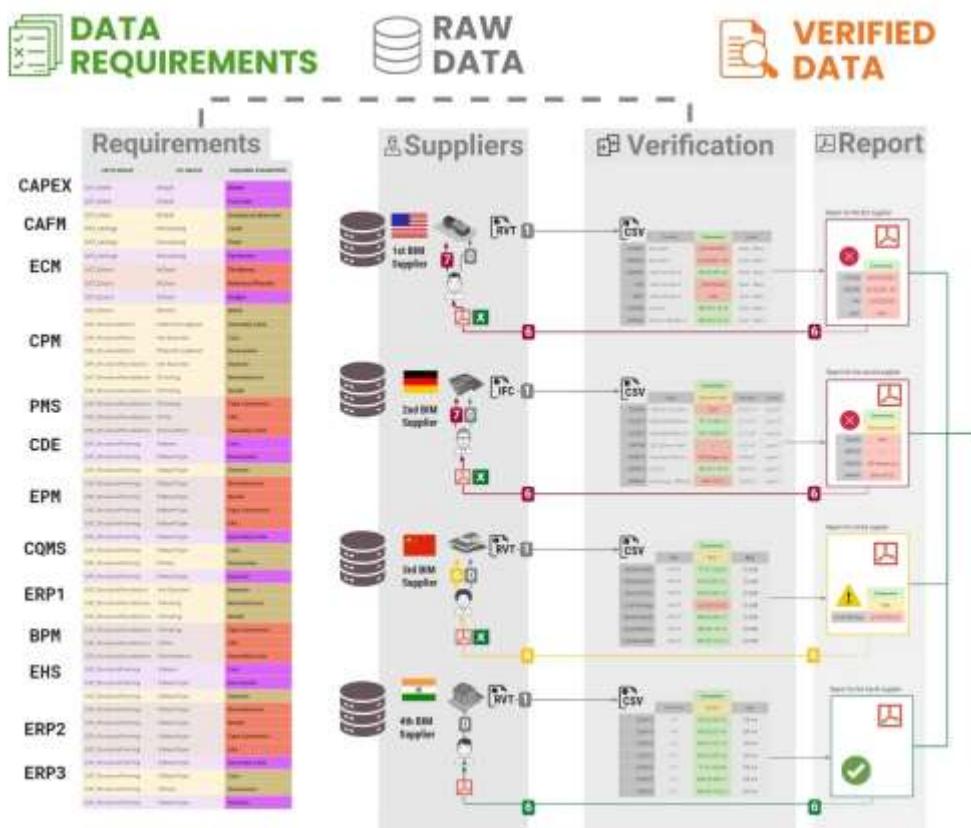
Proces validace dat CAD zahrnuje extrakci dat (fáze ETL Extract) z různých uzavřených (RVT, DWG, DGN, NWS atd.) nebo otevřených polostrukturovaných a parametrických formátů (IFC, CPXML, USD.) nebo otevřených polostrukturovaných a parametrických formátů (IFC, CPXML, USD), v nichž lze na jednotlivé atributy a jejich hodnoty aplikovat tabulky pravidel (fáze Transformace) pomocí regulárních výrazů RegEx (obr. 7.3-8), což je postup, který jsme podrobně rozebrali ve čtvrtém díle knihy.

Vytvoření chybové zprávy ve formátu PDF o úspěšně validovaných záznamech by mělo být doplněno výstupem (krok Load) ve strukturovaných formátech, které zohledňují pouze validované entity, jež lze použít pro další procesy.



Obrázek 7.3-8 Proces validace dat od poskytovatelů projektových dat po závěrečnou zprávu validovanou pomocí regulárních výrazů.

Automatizace validace dat ze systémů CAD (BIM) se strukturovanými požadavky a streamováním nových dat, která jsou zpracovávána prostřednictvím ETL-Pipelines (obrázek 7.3-9), snižuje potřebu manuální účasti na procesu validace (jednotlivé procesy validace a požadavků na data byly popsány v předchozích kapitolách).



Obr. 7.3-9 Automatizace validace dat prostřednictvím ETL zjednodušíuje řízení stavebních projektů tím, že urychluje procesy.

Ověřování modelů dodaných dodavateli a specialisty na CAD (BIM) může tradičně trvat dny až týdny.

Se zavedením automatizovaných procesů ETL se však tato doba může zkrátit na několik minut. V typické situaci zhotovitel uvádí: "*Model je validován a vyhovuje požadavkům.*" Tímto prohlášením začíná řetězec ověřování tvrzení zhotovitele o kvalitě dat:

- ☛ Projektový manažer - "Zhotovitel prohlašuje: '*Model byl otestován, vše je v pořádku.*'"
- ☛ Správce dat - Ověření načtení:
 - Jednoduchý skript v programu Pandas zjistí porušení během několika sekund. Automatizace odstraňuje spory:
 - Kategorie: OST_StructuralColumns, Parametr: FireRating IS NULL.
 - Vytvoření seznamu ID porušení → export do Excelu/PDF.

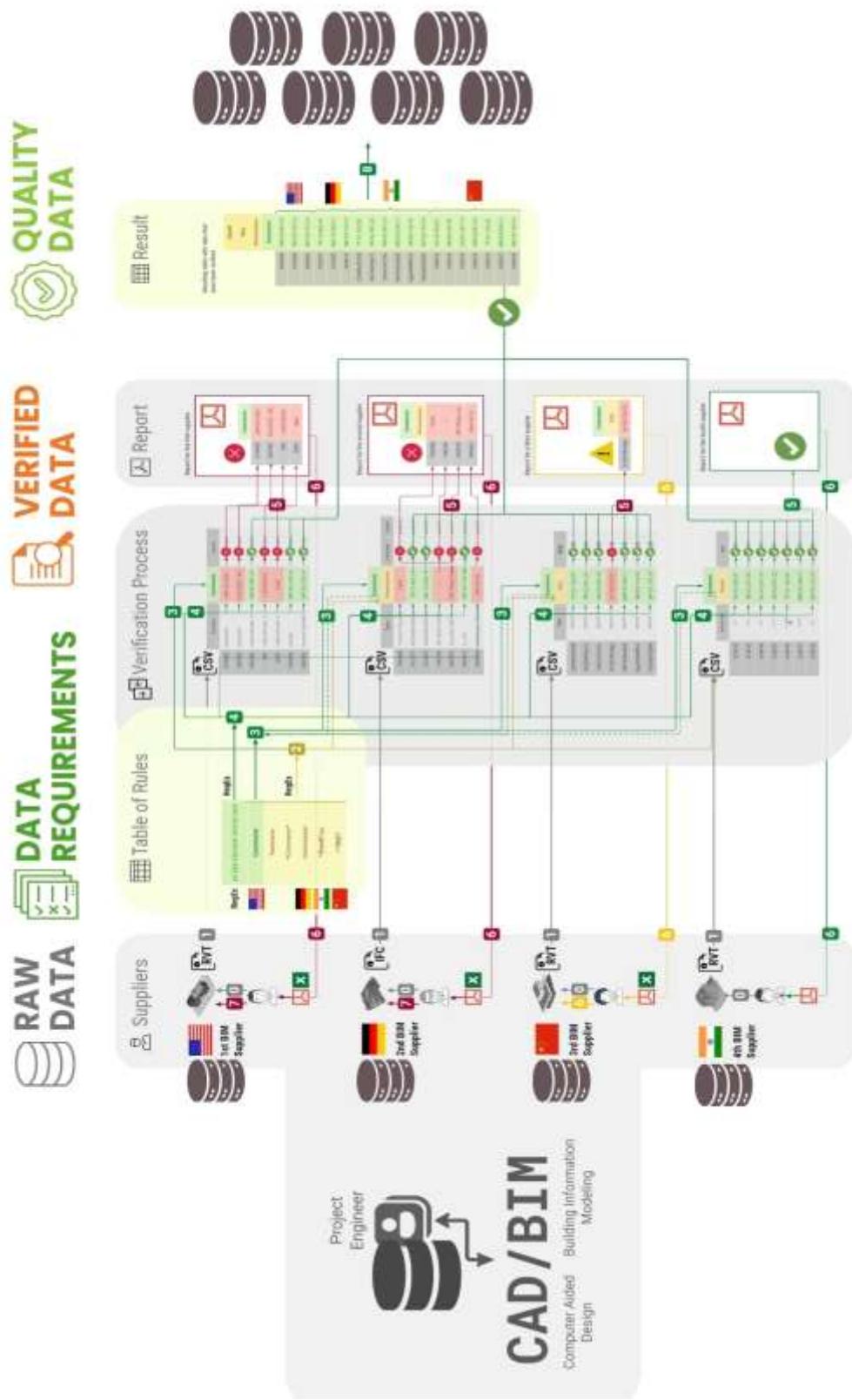
Jednoduchý skript v programu Pandas zjistí narušení během několika sekund:

```
df = model_data[model_data["Category"] == "OST_StructuralColumns"] # Filtrování
issues = df[df["FireRating"].isnull()] # Prázdné hodnoty
issues[["ElementID"]].to_excel("fire_rating_issues.xlsx") # Export IDs
```

- ☛ Správce dat projektovému manažerovi - "*Kontrola na ukázala, že 18 sloupců nemá vyplněný parametr FireRating.*"
- ☛ Projektový manažer zhotoviteli - "*Model se vrací k revizi: parametr FireRating je povinný, bez něj není možná akceptace.*"

Výsledkem je, že model CAD neprochází validací, automatizace eliminuje spory a zadavatel téměř okamžitě obdrží strukturovanou zprávu se seznamem ID problematických prvků. Tímto způsobem se proces validace stává transparentním, opakovatelným a chráněným před lidskou chybou (obr. 7.3-10).

Tento přístup mění proces validace dat na inženýrskou funkci namísto manuální kontroly kvality. Nejenže se tím zvyšuje produktivita, ale také je možné použít stejnou logiku na všechny projekty společnosti, což umožnuje komplexní digitální transformaci procesů od návrhu až po provoz.



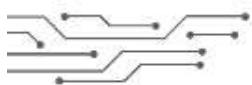
Obr. 7.3-10 Automatizace kontroly atributů prvků eliminuje lidské chyby a snižuje pravděpodobnost chyb.

Pomocí automatizovaných potrubí (obr. 7.3-10) mohou uživatelé systému, kteří očekávají kvalitní data ze systémů CAD- (BIM-), okamžitě získat potřebná výstupní data - tabulky, dokumenty, obrázky - a rychle je začlenit do svých pracovních úkolů.

Automatizace řízení, zpracování a analýzy je hnací silou změny přístupu k řízení stavebních projektů, zejména interoperability různých systémů, bez použití složitých a drahých modulárních proprietárních systémů nebo uzavřených dodavatelských řešení.

Zatímco pojmy a marketingové zkratky přicházejí a odcházejí, samotné procesy validace datových požadavků zůstanou navždy nedílnou součástí obchodních procesů. Namísto vytváření dalších a dalších specializovaných formátů a standardů by se stavebnictví mělo poohlédnout po nástrojích, které již prokázaly svou účinnost v jiných odvětvích. V současné době existují výkonné platformy pro automatizaci zpracování dat a integraci procesů, které firmám umožňují výrazně zkrátit čas strávený rutinními operacemi a minimalizovat chyby v Extract, Transform a Load.

Jedním z populárních příkladů řešení pro automatizaci a orchestraci procesů ETL je Apache Airflow, který umožňuje organizovat složité výpočetní procesy a spravovat potrubí ETL. Vedle Airflow se aktivně používají i další podobná řešení, jako je Apache NiFi pro směrování a streamování dat a n8n pro automatizaci podnikových procesů.



KAPITOLA 7.4.

ORCHESTRACE ETL A PRACOVNÍCH POSTUPŮ: PRAKTIČKÁ ŘEŠENÍ

DAG a Apache Airflow: automatizace a orchestrace pracovních postupů

Apache Airflow je bezplatná platforma s otevřeným zdrojovým kódem, určená k automatizaci, orchestraci a monitorování pracovních postupů (ETL -konvejerů).

Práce s velkým množstvím dat je nutná každý den:

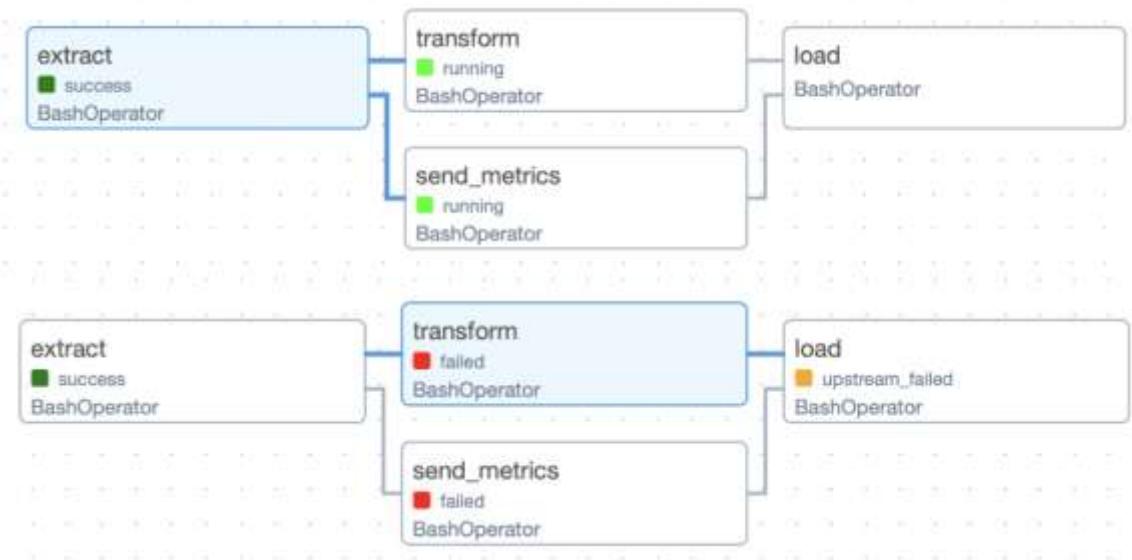
- Stahování souborů z různých zdrojů - Výpis (například od dodavatelů nebo zákazníků).
- Transformovat tato data do požadovaného formátu - Transformovat (strukturovat, vyčistit a ověřit).
- odesílání výsledků k ověření a vytváření zpráv - načítání (nahrávání do požadovaných systémů, dokumentů, databází nebo řídicích panelů).

Ruční provádění těchto procesů ETL vyžaduje značný čas a vede k riziku lidské chyby. Změna zdroje dat nebo selhání v některém z kroků může způsobit zpoždění a nesprávné výsledky.

Automatizační nástroje, jako je Apache Airflow, umožňují vytvořit spolehlivý ETL -převodník, minimalizovat chyby, zkrátit dobu zpracování a zajistit správnost dat v každé fázi. Základem Apache Airflow je koncept DAG (Directed Acyclic Graph) - směrovaný acyklický graf, ve kterém je každá úloha (operátor) spojena s dalšími závislostmi a vykonává se striktně v určeném pořadí. DAG eliminuje cykly, což poskytuje logickou a předvídatelnou strukturu provádění úloh.

Airflow se stará o orchestraci - spravuje závislosti mezi úlohami, řídí plány provádění, sleduje stav a automaticky reaguje na selhání. Tento přístup minimalizuje manuální zásahy a zajistuje spolehlivost celého procesu.

Orchestrátor úloh je nástroj nebo systém určený ke správě a řízení provádění úloh v komplexních výpočetních a informačních prostředích. Usnadňuje proces nasazování, automatizace a řízení provádění úloh s cílem zvýšit výkon a optimalizovat zdroje.



Obr. 7.4-1 Apache Airflow poskytuje uživatelsky přívětivé rozhraní, ve kterém lze vizualizovat DAG -ETL, zobrazit protokoly o provádění, stav spuštění úlohy a další.

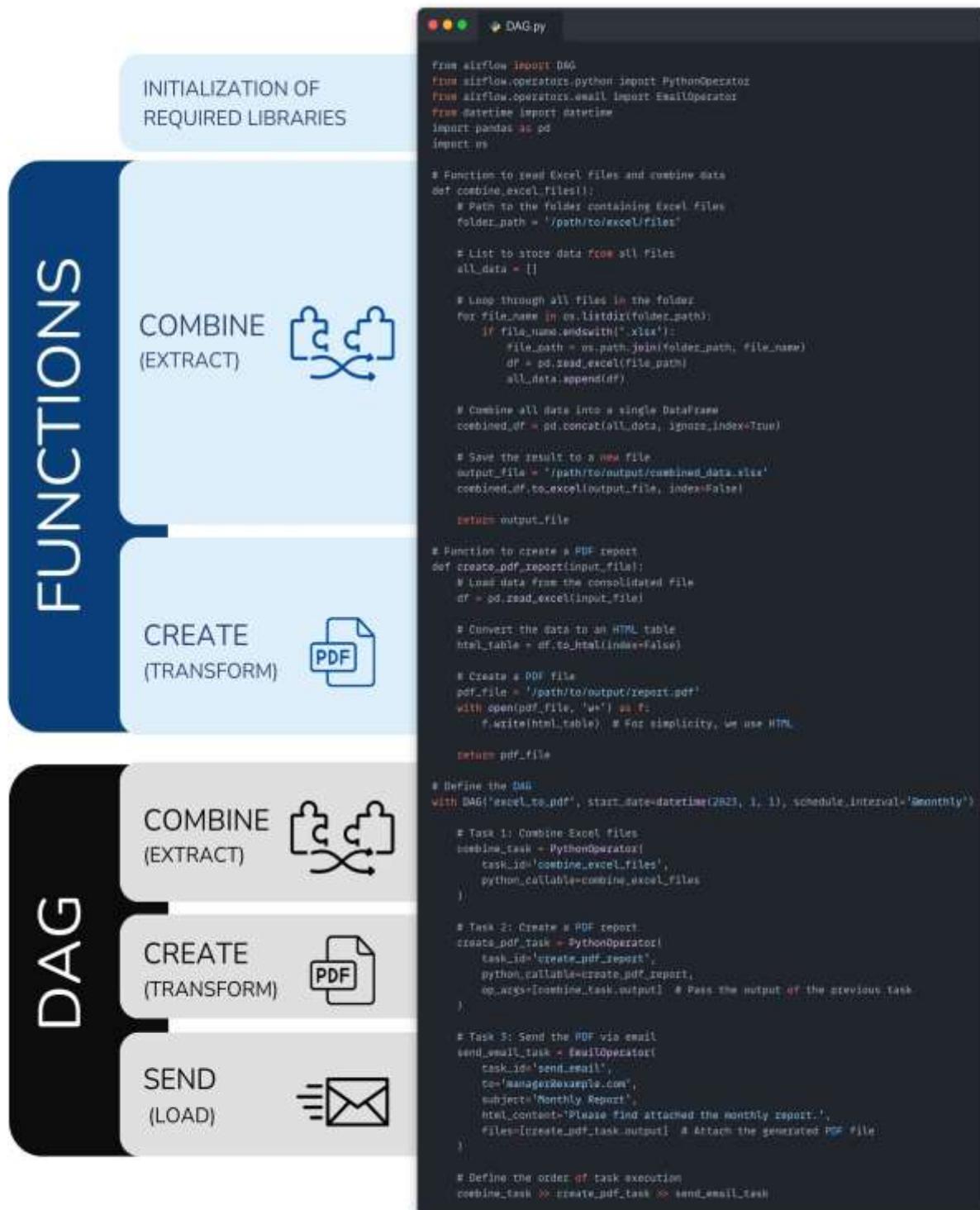
Airflow se široce používá pro orchestraci a automatizaci distribuovaných výpočtů, zpracování dat, řízení procesů ETL (Extract, Transform, Load), plánování úloh a dalších datových scénářů. Ve výchozím nastavení používá Apache Airflow jako databázi SQLite.

Příklad jednoduché DAG, podobně jako ETL, se skládá z úloh - Extract, Transform a Load. V grafu, který se ovládá přes uživatelské rozhraní (obr. 7.4-1), je definováno pořadí provádění úloh (fragmentů kódu): například nejprve se provede extrakce, pak transformace (a odeslání metrik) a práci dokončí úloha load. Jakmile jsou všechny úlohy dokončeny, je proces načítání dat považován za úspěšný.

Apache Airflow: praktická aplikace automatizaci ETL

Apache Airflow je široce používán k organizaci komplexních procesů zpracování dat a umožňuje vytvářet flexibilní ETL -konvejery. Apache Airflow lze spouštět buď přes webové rozhraní, nebo programově prostřednictvím kódu v jazyce Python (obr. 7.4-2). Ve webovém rozhraní (obr. 7.4-3) mohou správci a vývojáři vizuálně sledovat DAG, spouštět úlohy a analyzovat výsledky provádění.

Pomocí DAG, můžete nastavit jasnou posloupnost úloh, spravovat závislosti mezi nimi a automaticky reagovat na změny ve zdrojových datech. Uvedme si příklad použití nástroje Airflow k automatizaci zpracování hlášení (obr. 7.4-2).



Obr. 7.4-2 Koncept ETL -převodník pro zpracování dat pomocí Apache Airflow.

Tento příklad (obrázek 7.4-2) uvažuje DAG, která provádí klíčové úlohy v rámci ETL -převodníku:

ETL (EXTRACT, TRANSFORM, LOAD) PIPELINE

■ Čtení souborů aplikace Excel (Výpis):

- Sekvenční procházení všech souborů v daném adresáři.
- Čtení dat z každého souboru pomocí knihovny pandas.
- Sloučení všech dat do jednoho datového rámce.

■ Vytvořit PDF -dokument (Transform):

- Transformujte sloučený DataFrame do HTML -tabulky.
- Uložte tabulku jako PDF (v ukázkové verzi - prostřednictvím HTML).

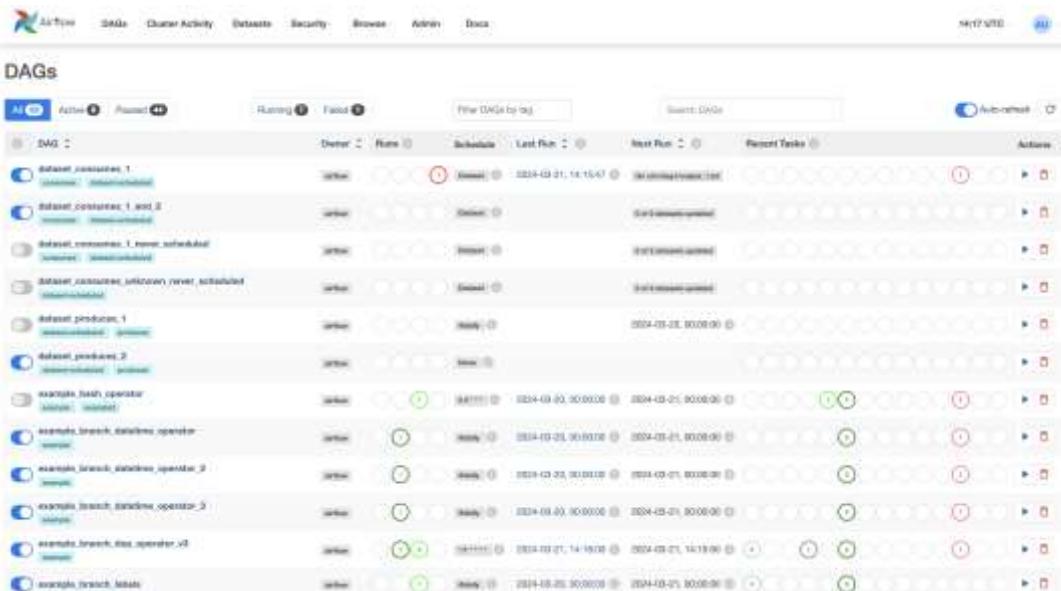
■ Odeslání zprávy e-mailem (Load):

- Použijte EmailOperator k odeslání PDF -dokumentu e-mailem.

■ Konfigurace skupiny DAG:

- Definování posloupnosti úloh: extrakce dat → generování sestavy → odeslání.
- Přiřazení plánu spouštění (@měsíčně - první den každého měsíce).

Automatický příklad ETL (obrázek 7.4-2) ukazuje, jak shromáždit data ze souborů aplikace Excel, vytvořit dokument PDF a odeslat jej e-mailem. Toto je jen jeden z mnoha možných případů použití aplikace Airflow. Tento příklad lze přizpůsobit libovolné konkrétní úloze, která zjednoduší a zautomatizuje zpracování dat.



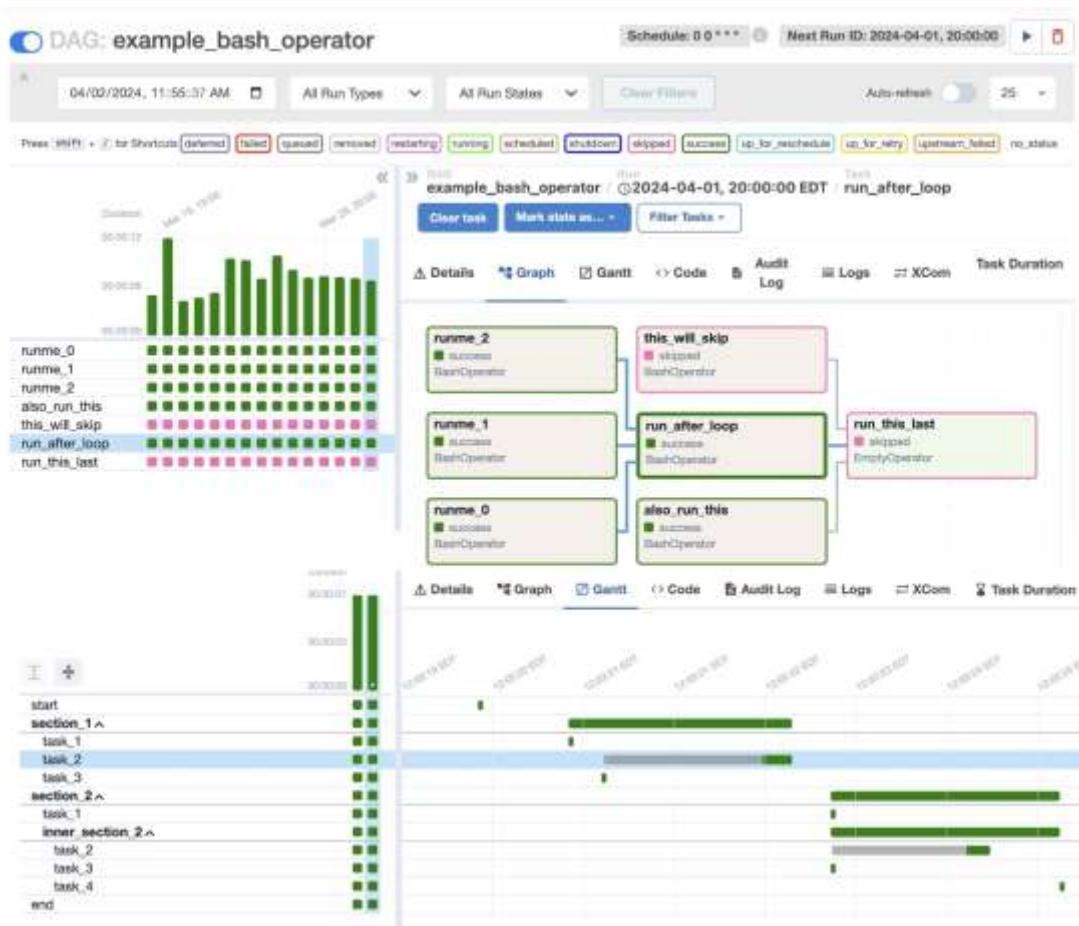
Obr. 7.4-3 Přehled všech DAG v prostředí s informacemi o posledních spuštěních.

Webové rozhraní Apache Airflow (obrázek 7.4-3) poskytuje komplexní vizuální prostředí pro správu pracovních postupů s daty. Zobrazuje DAG jako interaktivní grafy, kde uzly představují úlohy a hrany závislosti mezi nimi, což usnadňuje sledování složitých datových pracovních postupů. Rozhraní obsahuje ovládací panel s informacemi o stavu provádění úloh, historii běhu, podrobných protokolech a výkonnostních metrikách. Správci mohou ručně spouštět úlohy, restartovat neúspěšné operace, pozastavovat DAG a konfigurovat proměnné prostředí, a to vše prostřednictvím intuitivního uživatelského rozhraní.

Takovou architekturu lze doplnit o validaci dat, oznámení o stavu provádění, integraci s externími rozhraními API nebo databázemi. Airflow umožnuje flexibilně přizpůsobovat DAG: přidávat nové úlohy, měnit jejich pořadí, kombinovat řetězce - což z něj činí efektivní nástroj pro automatizaci komplexních procesů zpracování dat. Při spuštění DAG ve webovém rozhraní Airflow (obr. 7.4-3, obr. 7.4-4) můžete sledovat stav úloh. Systém používá barevnou indikaci:

- Zelená - úkol byl úspěšně dokončen.
- Žlutá - proces probíhá.
- Červená - chyba při provádění úkolu.

V případě selhání (např. chybějící soubor nebo poškozená datová struktura) systém automaticky zahájí odeslání oznámení.



Obrázek 7.4-4 Apache Airflow výrazně zjednodušíuje diagnostiku problémů, optimalizaci procesů a týmovou spolupráci na složitých pipelinech zpracování dat.

Aplikace Apache Airflow je pohodlná, protože automatizuje rutinní úkoly a eliminuje nutnost provádět je ručně. Zajišťuje spolehlivost tím, že monitoruje provádění procesů a poskytuje okamžité upozornění na chyby. Flexibilita systému umožňuje snadno přidávat nové úlohy nebo upravovat stávající a přizpůsobovat pracovní postupy měnícím se požadavkům.

Kromě aplikace Apache Airflow, existují podobné nástroje pro orchestraci pracovních postupů. Například open source a bezplatný Prefect (obrázek 7.3-5) nabízí jednodušší syntaxi a lépe se integruje s jazykem Python, Luigi, vyvinutý společností Spotify, poskytuje podobné funkce a dobře pracuje s velkými daty. Za zmínku stojí také Kronos a Dagster, které nabízejí moderní přístupy k budování Pipeline se zaměřením na modularitu a škálovatelnost. Výběr nástroje pro orchestraci úloh závisí na konkrétních potřebách projektu, ale všechny pomáhají automatizovat složité datové procesy ETL.

Za zmínku stojí zejména Apache NiFi, open source platforma, určená pro streamování a směrování dat. Na rozdíl od platformy Airflow, která se zaměřuje na dávkové zpracování a správu závislostí, se NiFi soustředí na transformaci dat v reálném čase za běhu a flexibilní směrování mezi systémy.

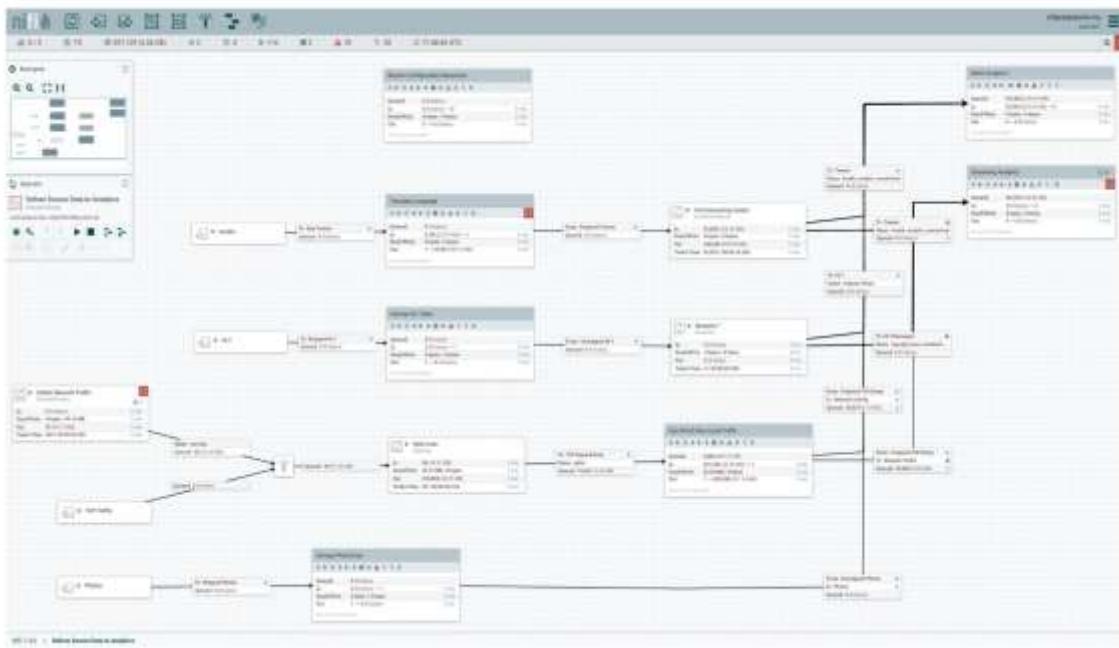
Apache NiFi pro směrování a konverzi dat

Apache NiFi je výkonná open source platforma, určená k automatizaci datových toků mezi různými systémy. Původně byla vyvinuta v roce 2006 americkou Národní bezpečnostní agenturou (NSA) pod názvem "Niagara Files" pro interní použití. V roce 2014 byl projekt otevřen a převeden na nadaci Apache Software Foundation, čímž se stal součástí jejích iniciativ pro přenos technologií [141].

Aplikace Apache NiFi je určena ke sběru, zpracování a přenosu dat v reálném čase. Na rozdíl od Airflow, který pracuje s dávkovými úlohami a vyžaduje přesně definované plány, NiFi pracuje v režimu proudového zpracování, což umožňuje nepřetržitý přenos dat mezi různými službami.

Apache NiFi je ideální pro integraci se zařízeními IoT, stavebními senzory, monitorovacími systémy a např. pro streamované ověřování formátů CAD na serveru, kde může být vyžadována okamžitá reakce na změny dat.

Díky vestavěným nástrojům pro filtrování, transformaci a směrování umožňuje NiFi standardizaci dat (Transform) před jejich přenosem (Load) do úložišť nebo analytických systémů. Jednou z jeho hlavních výhod je vestavěná podpora zabezpečení a řízení přístupu, což z něj činí spolehlivé řešení pro nakládání s citlivými informacemi.



Obrázek 7.4-5 Grafické znázornění toku dat v rozhraní Apache NiFi.

Apache NiFi efektivně zpracovává úlohy streamování, filtrování a směrování dat v reálném čase. Je ideální pro technicky náročné scénáře, kde je důležitý stabilní přenos informací mezi systémy a vysoká propustnost.

Pokud je však hlavním cílem integrace různých služeb, automatizace rutinních operací a rychlé nastavení pracovních postupů bez hlubších znalostí programování, jsou žádaná řešení s nízkým vstupním prahem a maximální flexibilitou. Jedním z takových nástrojů je n8n - platforma třídy Low-Code /No-Code zaměřená na automatizaci podnikání a vizuální orchestraci procesů.

n8n Low-Code, No-Code orchestrace procesů.

n8n je Open Source Low-Code / No-Code platforma pro vytváření automatizovaných pracovních postupů, která se vyznačuje snadným použitím, flexibilitou a schopností rychlé integrace s širokou škálou externích služeb.

No-Code je metoda vytváření digitálních produktů bez psaní kódu. Všechny prvky procesu - od logiky až po rozhraní - jsou realizovány výhradně pomocí vizuálních nástrojů. Platformy No-Code jsou určeny uživatelům bez technického vzdělání a umožňují rychle vytvářet automatizace, formuláře, integrace a webové aplikace. Příklad: uživatel nastaví automatické zasílání oznámení nebo integraci s tabulkami Google prostřednictvím rozhraní drag-and-drop bez znalosti programování.

Díky možnostem nasazení s otevřeným zdrojovým kódem a lokálnímu nasazení poskytuje n8n v

procesech automatizace a tvorby ETL Pipelines společnostenem úplnou kontrolu nad jejich daty a zároveň zajišťuje bezpečnost a nezávislost na poskytovatelích cloudových služeb.

Na rozdíl od aplikace Apache Airflow, která je orientována na výpočetní úlohy s pevnou orchesterací a vyžaduje znalost jazyka Python, n8n poskytuje vizuální editor, který umožňuje psaní skriptů bez nutnosti znalosti programovacích jazyků (obrázek 7.4-6). Přestože jeho rozhraní umožňuje vytvářet automatizované procesy bez nutnosti psát kód (No-Code), ve složitějších scénářích mohou uživatelé přidávat vlastní funkce v jazycích JavaScript a Python - a rozšiřovat tak jeho možnosti (Low-Code).

Low-Code je přístup k vývoji softwaru, při kterém se základní logika aplikace nebo procesu vytváří pomocí grafického rozhraní a vizuálních prvků a programový kód se používá pouze k přizpůsobení nebo rozšíření funkčnosti. Platformy Low-Code umožňují výrazně urychlit vývoj řešení zapojením nejen programátorů, ale i podnikových uživatelů se základními technickými dovednostmi. Příklad: uživatel může sestavit obchodní proces z připravených bloků a v případě potřeby přidat vlastní skript v jazyce JavaScript nebo Python.

Ačkoli je n8n umístěn jako platforma s nízkým vstupním prahem, je třeba mít základní znalosti programování, porozumění webovým technologiím a dovednosti v práci s API. Flexibilita systému umožňuje jeho přizpůsobení široké škále úloh - od automatizovaného zpracování dat až po integraci s messengery, zařízeními internetu věcí a cloudovými službami.

Hlavní funkce a výhody používání n8n:

- **Otevřený zdrojový kód** a možnosti lokálního nasazení zajišťují úplnou kontrolu nad daty, dodržování bezpečnostních předpisů a nezávislost na poskytovatelích cloudových služeb.
- **Integrace s více než 330 službami** včetně CRM, ERP, elektronického obchodu, cloudových platforem, messengerů a databází.
- **Flexibilita scénářů**: od jednoduchých oznámení až po komplexní řetězce se zpracováním požadavků API, logikou rozhodování a připojením služeb AI.
- **Podpora pro JavaScript a Python**: uživatelé mohou podle potřeby vkládat vlastní kód a rozšiřovat tak možnosti automatizace.
- **Intuitivní vizuální rozhraní**: umožňuje rychlou konfiguraci a vizualizaci všech kroků procesu.

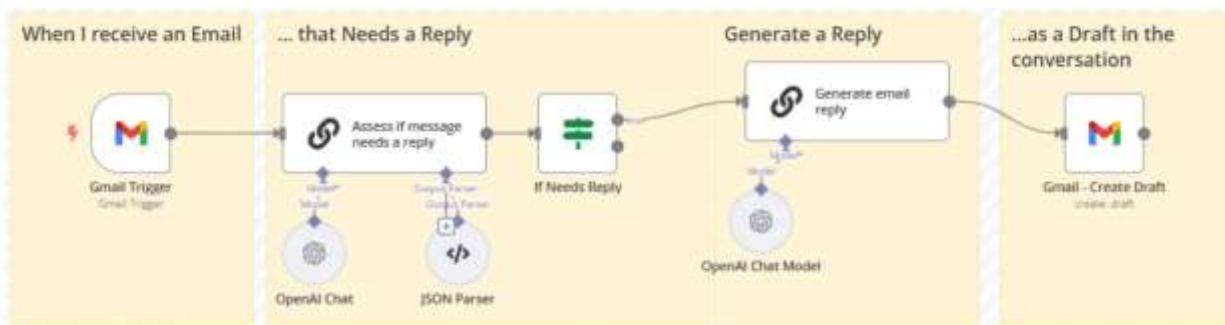
Platformy třídy Low-Code poskytují nástroje pro vytváření digitálních řešení s minimem kódu, takže jsou ideální pro týmy, které nemají hluboké technické znalosti, ale potřebují automatizovat procesy.

Ve stavebnictví lze n8n využít k automatizaci řady procesů, jako je integrace se systémy pro řízení projektů, kontrola toku, psaní zpráv a dopisů z , automatická aktualizace údajů o zásobách materiálu, zasílání oznámení o stavu úkolů týmům a další. Přizpůsobený Pipeline v n8n může několikanásobně snížit počet manuálních operací, snížit pravděpodobnost chyb a urychlit rozhodování o realizaci projektu.

Můžete si vybrat z téměř dvou tisíc hotových, bezplatných a otevřených n8n Pipeline, které jsou k dispozici na adrese: n8n.io/workflows, a automatizovat tak pracovní postupy ve stavebnictví i osobní úkoly a omezit rutinní operace.

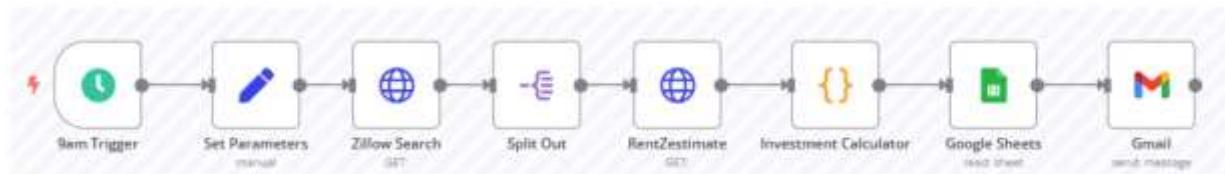
Vezměte si jednu z připravených šablon Pipeline, která je zdarma k dispozici na adrese n8n.io [142] a která automaticky vytváří návrhy odpovědí v aplikaci Gmail (obr. 7.4-6), což pomáhá uživatelům, kteří dostávají velké množství e-mailů nebo mají potíže s tvorbou odpovědí.

Tato šablona n8n "Gmail AI Auto-Responder: Create draft responses to incoming emails" (Obrázek 7.4-6) analyzuje příchozí e-maily pomocí LLM z ChatGPT, určí, zda je třeba odpovědět, vytvoří koncept z ChatGPT a převede text do HTML a přidá jej do řetězce zpráv v Gmailu. Tím se e-mail automaticky neodešle, což umožňuje ruční úpravu a schválení odpovědi. Nastavení trvá asi 10 minut a zahrnuje konfiguraci OAuth rozhraní Gmail API a integraci rozhraní OpenAI API. Výsledkem je pohodlné a bezplatné řešení pro automatizaci běžné e-mailové komunikace bez ztráty kontroly nad obsahem e-mailů.



Obr. 7.4-6 Automatizovaný proces generování e-mailové odpovědi pomocí n8n.

Dalším příkladem automatizace pomocí n8n je vyhledávání výhodných nabídek na trhu s nemovitostmi [143]. N8n Pipeline "Automate daily real deals using Zillow API, Google Sheets and Gmail", shromažďuje denně relevantní nabídky odpovídající zadaným kritériím pomocí rozhraní Zillow API. Automaticky počítá klíčové investiční ukazatele (Cash on Cash ROI, Monthly Cash Flow, Down Payment), aktualizuje Google Sheets a posílá souhrnnou zprávu na e-mail (obrázek 7.4-7), což investorům umožňuje ušetřit čas a rychle reagovat na nejlepší nabídky.



Obr. 7.4-7 Automatizovaný proces hodnocení investiční atraktivity nemovitostí.

Díky své flexibilitě a rozšiřitelnosti je n8n cenným nástrojem pro společnosti, které se snaží digitálně transformovat a stát se konkurenceschopnějšími na trhu pomocí relativně jednoduchých a bezplatných nástrojů s otevřeným zdrojovým kódem.

Nástroje jako Apache NiFi, Airflow a n8n si lze představit jako tři vrstvy zpracování dat (obrázek 7.4-8). NiFi řídí tok dat, zajišťuje jejich doručení a transformaci, Airflow orcheSTRUje provádění úloh agregací dat do zpracovatelských pipeline a n8n automatizuje integraci s externími službami a řídí obchodní logiku.

	The main task	Approach
Apache NiFi	Streaming and data transformation	Real-time stream processing
Apache Airflow	Task orchestration, ETL pipelines	Batch planning, DAG processes
n8n	Integration, automation of business logic	Low-code visual orchestration

Obr. 7.4-8 Apache Airflow, Apache NiFi a n8n lze považovat za tři vzájemně se doplňující vrstvy moderní architektury správy dat.

Tyto bezplatné nástroje s otevřeným zdrojovým kódem mohou společně tvořit příklad efektivního ekosystému pro správu dat a procesů ve stavebnictví a umožnit společnostem využívat informace pro rozhodování a automatizaci procesů.

Další kroky: přechod od manuálních operací k řešením založeným na analytice

Dnešní stavební společnosti pracují v prostředí vysoké nejistoty: měnící se ceny materiálů, zpožděné dodávky, nedostatek pracovních sil a napjaté termíny projektů. Využití analytických panelů, ETL - konvejerů a systémů BI pomáhá společnostem rychle identifikovat problémové oblasti, vyhodnocovat efektivitu zdrojů a předvídat změny dříve, než povedou k finančním ztrátám.

Pro shrnutí této části je vhodné zdůraznit hlavní praktické kroky, které vám pomohou aplikovat probírané technologie v každodenních úkolech:

- Implementace vizualizací dat a analytických panelů
 - Zvládnout proces vytváření řídicích panelů pro sledování klíčových ukazatelů výkonnosti (KPI).
 - Používejte nástroje pro vizualizaci dat (Power BI, Tableau, Matplotlib, Plotly).
- Automatizace zpracování dat prostřednictvím ETL -procesů
 - Nastavení automatického sběru dat z různých zdrojů (dokumentace, tabulky, CAD) prostřednictvím procesů ETL.
 - Organizovat transformaci dat (např. kontrolu regulárních výrazů nebo výpočet) pomocí

skriptů Pythonu.

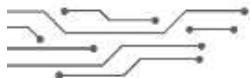
- Vyzkoušejte nastavení automatického vytváření zpráv ve formátu PDF (nebo DOC) pomocí knihovny FPDF s využitím dat ze souborů aplikace Excel nebo extrahováním informací z jiných dokumentů PDF.

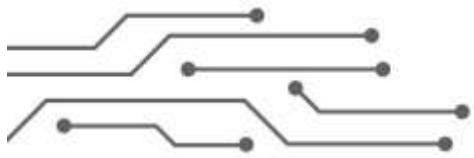
■ Použití jazykových modelů (LLM) pro automatizaci

- použití rozsáhlých jazykových modelů (LLM) pro generování kódu, který pomáhá extrahovat a analyzovat data z nestrukturovaných dokumentů.
- Seznamte se s automatizačním nástrojem n8n a prozkoumejte připravené šablony a případové studie na jejich webových stránkách. Určete, které procesy z vaší práce lze plně automatizovat pomocí přístupu No-Code/Low-Code.

Analytický přístup k datům a automatizaci procesů nejen zkracuje čas strávený rutinními operacemi, ale také zlepšuje kvalitu rozhodování. Společnosti, které implementují vizuální analytické nástroje a ETL -převodníky, získávají možnost rychle reagovat na změny

Automatizace podnikových procesů pomocí nástrojů, jako jsou n8n, Airflow a NiFi, je pouze prvním krokem k digitální vyspělosti. Dalším krokem je kvalitní ukládání a správa samotných dat, která jsou základem automatizace. V 8. díle se podrobně podíváme na to, jak mohou stavební společnosti vybudovat udržitelnou architekturu ukládání dat a přejít od chaosu dokumentů a souborů různých formátů k centralizovaným platformám pro ukládání a analýzu.





VIII ČÁST

UKLÁDÁNÍ A SPRÁVA DAT VE STAVEBNICTVÍ

Část 8 se zabývá moderními technologiemi pro ukládání a správu dat ve stavebnictví. Analyzuje efektivní formáty pro práci s velkými objemy informací - od jednoduchých CSV a XLSX až po výkonnější Apache Parquet a ORC s podrobným porovnáním jejich možností a omezení. Rozebírá koncepty datových skladů (DWH), datových jezer) a jejich hybridních řešení (Data Lakehouse), jakož i principy Data Governance) a Data Minimalism). Podrobně jsou popsány problémy datové bažiny) a strategie, jak chaosu v informačních systémech předcházet. Jsou představeny nové přístupy k práci s daty, včetně vektorových databází a jejich aplikace při konstrukci prostřednictvím konceptu Bounding Box. Tato část se také dotýká metodik DataOps a VectorOps jako nových standardů pro organizaci práce s daty.

KAPITOLA 8.1.

DATOVÁ INFRASTRUKTURA: OD ÚLOŽNÝCH FORMÁTŮ K DIGITÁLNÍM ÚLOŽIŠTÍM

Datové atomy: základ efektivní správy informací

Všechno ve vesmíru se skládá z nejmenších stavebních kamenů - atomů a molekul a postupem času se vše živé i neživé nevyhnutelně vrací do tohoto výchozího stavu. V přírodě tento proces probíhá s ohromující rychlostí, kterou se snažíme přenést na procesy řízené člověkem.

V lese se všechny živé organismy nakonec přemění na výživnou látku, která slouží jako základ pro nové rostliny. Tyto rostliny se zase stávají potravou pro nové živé organismy složené ze stejných atomů, které před miliony let stvořily vesmír.

Ve světě podnikání je také důležité rozložit složité, mnohovrstevnaté struktury na nejzákladnější, minimálně zpracované jednotky podobně jako atomy a molekuly v přírodě. Díky tomu lze atomy dat efektivně ukládat a spravovat, čímž se z nich stane bohatý, úrodný základ, který se stane klíčovým zdrojem pro růst analytiky a kvality rozhodování.



Obr. 8.1-1 Analýza a rozhodování je založeno na opakovaně používaných datech, která byla jednou zpracována a uložena.

Hudební skladby se skládají z tónů, které se spojují do složitých hudebních děl, zatímco slova vznikají z primitivní jednotky - písmene a zvuku. Ať už se jedná o přírodu, vědu, ekonomiku, umění nebo

technologií, svět vykazuje pozoruhodnou jednotu a harmonii ve své snaze o destrukci, strukturu, cyklus a tvorbu. Podobně jsou procesy v systémech kalkulace nákladů rozděleny na drobné strukturované jednotky - položky zdrojů - na úrovni kalkulací a harmonogramů. Tyto jednotky, podobně jako poznámky, se pak používají k tvorbě složitějších kalkulací a rozvrhů. Stejný princip využívají systémy počítačového projektování, v nichž se složité architektonické a inženýrské projekty sestavují ze základních prvků - jednotlivých prvků a knihovních komponent, z nichž se vytváří kompletní 3D -model projektu složité budovy nebo stavby.

Koncept cyklicity a struktury, který je vlastní přírodě a vědě, se odráží i v moderním světě dat. Stejně jako se v přírodě vše živé vrací k atomům a molekulám, tak i ve světě moderních nástrojů pro zpracování dat mají informace tendenci vracet se ke své nejprimitivnější podobě.

Nejmenší prvky s jejich konečnou nedělitelností jsou základními stavebními kameny podnikových procesů. Je důležité od počátku pečlivě zvažovat, jak tyto drobné stavební prvky z různých zdrojů shromažďovat, strukturovat (rozkládat na atomy) a ukládat. Organizace a ukládání dat není jen otázkou jejich rozkladu na jednotlivé složky. Stejně důležité je zajistit jejich integraci a strukturované uložení, aby bylo možné data snadno vyhledat, analyzovat a použít pro rozhodování, kdykoli je to potřeba.

Pro efektivní zpracování informací je třeba pečlivě zvolit formát a metody ukládání dat - stejně jako je třeba připravit půdu pro růst stromů. Datová úložiště musí být uspořádána tak, aby byla zajištěna vysoká kvalita a relevance informací a aby byla vyloučena nadbytečná nebo nerelevantní data. Čím lépe je tato "informační půda" strukturována, tím rychleji a přesněji mohou uživatelé najít správná data a řešit analytické problémy.

Ukládání informací: soubory nebo data

Datové sklady umožňují společnostem shromažďovat a kombinovat informace z různých systémů a vytvářet tak jediné centrum pro další analýzy. Shromázděná historická data umožňují nejen hlubší analýzu procesů, ale také identifikaci vzorců, které mohou ovlivnit výkonnost podniku.

Řekněme, že společnost pracuje na několika projektech současně. Inženýr chce zjistit, kolik betonu bylo vylito a kolik ho ještě zbývá nakoupit. Při tradičním přístupu by musel ručně prohledat server a otevřít několik tabulek odhadů, porovnat je s certifikáty provedených prací a zkontoval aktuální stav y zásob. To trvá hodiny nebo dokonce dny. Dokonce i s procesy ETL a automatickými skripty zůstává úloha polomanuální: inženýr musí stále ručně zadávat cesty ke složkám nebo konkrétním souborům na serveru. To sniže celkový efekt automatizace, protože nadále zabírá cenný pracovní čas.

Při přechodu na správu dat získá inženýr místo práce se souborovým systémem serveru přístup k jednotné struktuře úložiště, kde jsou informace aktualizovány v reálném čase. Jediný dotaz - ve formě kódu, SQL -dotazu nebo dokonce volání LLM -agenta - může okamžitě poskytnout přesné údaje o aktuálních zůstatcích, provedených pracích a nadcházejících dodávkách, pokud byla data předem připravena a sdružena v datovém skladu, kde není třeba bloudit po složkách, otevírat desítky souborů a ručně porovnávat hodnoty.

Stavební firmy dlouhou dobu používaly PDF -dokumenty, DWG -výkresy, RVT -modely a stovky a tisíce excelových -tabulky a další nesourodé formáty, které jsou uloženy ve specifických složkách na firemních serverech, což ztěžuje vyhledávání informací, jejich kontrolu a analýzu. Výsledkem je, že soubory, které po dokončení projektů zůstanou, se nejčastěji přesouvají zpět na server do archivních složek, které se v budoucnu prakticky nepoužívají. Takové tradiční ukládání dat založené na souborech ztrácí s rostoucím tokem dat na významu kvůli své náchylnosti k lidským chybám.

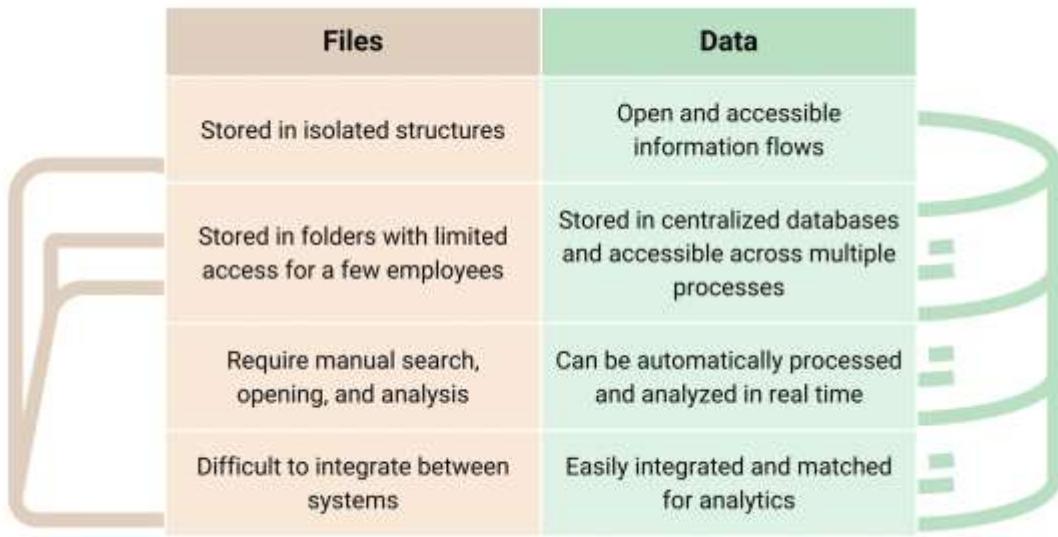
Soubor je pouze izolovaný kontejner, ve kterém jsou uložena data. Soubory jsou vytvářeny pro lidi, nikoli pro systémy, takže vyžadují ruční otevřání, čtení a interpretaci. Příkladem může být tabulka v aplikaci Excel -tabulka, dokument ve formátu PDF -dokument nebo výkres CAD -výkres, který je třeba speciálně otevřít v určitém nástroji, abyste získali přístup k požadovaným informacím. Bez strukturovaného vyhledávání a zpracování zůstanou informace v něm obsažené nevyužity.

Data jsou zase strojově čitelné informace, které se automaticky propojují, aktualizují a analyzují. V jednotném datovém skladu (např. databázi, DWH nebo Data Lake) jsou informace reprezentovány ve formě tabulek, záznamů a vztahů. To umožňuje jednotné ukládání, automatizované dotazy, analýzy hodnot a reporting v reálném čase.

Použití dat namísto souborů (obr. 8.1-1) umožňuje eliminovat proces ručního vyhledávání a sjednotit zpracování. Společnosti, které již takový přístup zavedly, získávají konkurenční výhodu díky rychlosti přístupu k informacím a schopnosti rychle je integrovat do podnikových procesů.

Přechod od používání souborů k datům je nevyhnutelnou změnou, která ovlivní budoucnost stavebnictví.

Každá společnost ve stavebnictví bude stát před klíčovou volbou: pokračovat v ukládání informací do různorodých souborů a sil, které musí čist lidé pomocí speciálních programů, nebo je v prvních fázích zpracování transformovat do strukturovaných dat a vytvořit jednotný integrovaný digitální základ pro automatizované řízení projektů.



Obrázek 8.1-1 Vývoj toku informací: od izolovaných souborů k integrovaným datům.

S prudkým nárůstem objemu informací jsou tradiční metody ukládání a zpracování souborů stále méně efektivní. Ve stavebnictví, stejně jako v jiných odvětvích, již nestačí spoléhat se na různorodé složky s různými formáty souborů nebo na nepropojené databáze.

Společnosti, které si chtějí udržet konkurenceschopnost v digitálním věku, budou nevyhnutelně přecházet na integrované digitální platformy, využívat technologie big data a automatizované analytické systémy.

Přechod od ukládání souborů k pracovním postupům založeným na datech bude vyžadovat přehodnocení přístupů ke správě informací a vědomou volbu formátů vhodných pro další integraci do centralizovaných úložišť. Tato volba určí, jak efektivně lze data zpracovávat, jak rychle k nim lze přistupovat a jak snadno je lze integrovat do digitálních procesů společnosti.

Ukládání velkých objemů dat: Analýza oblíbených formátů a jejich efektivity

Formáty úložišť hrají klíčovou roli při škálovatelnosti, spolehlivosti a výkonu analytické infrastruktury. K analýze a zpracování dat - například filtrování, seskupování a agregaci - jsme v našich příkladech použili Pandas DataFrame - oblíbenou strukturu pro práci s daty v paměti RAM.

Pandas DataFrame však nemá vlastní formát pro ukládání, takže po dokončení zpracování se data exportují do některého z externích formátů - nejčastěji CSV nebo XLSX. Tyto tabulkové formáty se snadno vyměňují a jsou kompatibilní s většinou externích systémů, ale mají řadu omezení: nízkou efektivitu ukládání, nedostatečnou kompresi a špatnou podporu verzování:

- **CSV** (Comma-Separated Values): jednoduchý textový formát široce podporovaný různými platformami a nástroji. Snadno se používá, ale nepodporuje složité datové typy a kompresi.
- **XLSX** (Excel Open XML Spreadsheet): formát souboru aplikace Microsoft Excel, který podporuje sofistikované funkce, jako jsou vzorce, grafy a stylování. Je sice užitečný pro

manuální analýzu a vizualizaci dat, ale není optimalizován pro rozsáhlé zpracování dat.

Kromě oblíbených tabulkových formátů XLSX a CSV, existuje několik populárních formátů pro efektivní ukládání strukturovaných dat (obrázek 8.1-2), z nichž každý má jedinečné výhody v závislosti na konkrétních požadavcích na ukládání a analýzu dat:

- **Apache Parquet:** formát sloupcového ukládání dat optimalizovaný pro použití v systémech pro analýzu dat. Nabízí účinná schémata komprese a kódování dat, takže je ideální pro komplexní datové struktury a zpracování velkých objemů dat.
- **Apache ORC (Optimised Row Columnar):** Podobně jako Parquet, ORC poskytuje vysokou kompresi a efektivní ukládání dat. Je optimalizován pro náročné operace čtení a je vhodný pro ukládání datových jezer.
- **JSON (JavaScript Object Notation):** JSON sice není tak efektivní z hlediska ukládání dat ve srovnání s binárními formáty, jako je Parquet nebo ORC, ale je velmi přístupný a snadno se s ním pracuje, takže je ideální pro skripty, u nichž je důležitá čitelnost a kompatibilita s webem.
- **Feather:** rychlý, lehký a snadno použitelný binární sloupcový formát pro ukládání dat orientovaný na analýzu. Je navržen pro efektivní přenos dat mezi jazyky Python (Pandas) a R, takže je vynikající volbou pro projekty zahrnující tato programovací prostředí.
- **HDF5 (Hierarchical datový formát verze 5):** určený pro ukládání a organizaci velkého množství dat. Podporuje širokou škálu datových typů a je vhodný pro práci se složitými soubory dat. HDF5 je oblíbený zejména ve vědeckých výpočtech díky své schopnosti efektivně ukládat a zpřístupňovat velké soubory dat.

	XLSX	CSV	Apache Parquet	HDF5	Pandas DataFrame	
	Storage	Tabular	Tabular	Columnar	Hierarchical	Tabular
	Usage	Office tasks, data presentation	Simple data exchange	Big data, analytics	Scientific data, large volumes	Data analysis, manipulation
	Compression	Built-in	None	High	Built-in	None (in-memory)
	Performance	Low	Medium	High	High	High (memory dependent)
	Complexity	High (formatting, styles)	Low	Medium	Medium	Low
	Data Type Support	Limited	Very limited	Extended	Extended	Extended
	Scalability	Low	Low	High	High	Medium (memory limited)

Obr. 8.1-2 Srovnání datových formátů, které ukazuje hlavní rozdíly v aspektech ukládání a zpracování.

Pro provedení srovnávací analýzy formátů používaných ve fázi Load procesu ETL byla vytvořena tabulka zobrazující velikosti souborů a časy čtení (obr. 8.1-3). Ve studii byly použity soubory s identickými daty: tabulka obsahovala 10 000 řádků a 10 sloupců vyplněných náhodnými hodnotami.

Do studie jsou zahrnutý následující formáty ukládání: CSV, Parquet, XLSX a HDF5, jakož i jejich komprimované verze v archivech ZIP. Surová data byla generována pomocí knihovny NumPy a reprezentována jako struktura Pandas DataFrame. Proces testování se skládal z následujících kroků:

- Ukládání souborů: datový rámec se ukládá ve čtyřech různých formátech: CSV, Parquet, XLSX a HDF5. Každý formát má jedinečné vlastnosti ve způsobu ukládání dat, což ovlivňuje velikost souboru a rychlosť čtení.
- Komprese souborů ZIP: za účelem analýzy účinnosti standardní komprese byl každý soubor dále komprimován do archivu ZIP.
- Čtení souborů (ETL - Load): doba čtení byla měřena pro každý soubor po rozbalení ze ZIP. To umožňuje odhadnout rychlosť přístupu k datům po rozbalení z archivu.

Je důležité poznamenat, že Pandas DataFrame nebyl použit přímo při analýze velikosti nebo doby čtení, protože nepředstavuje samostatný formát úložiště. Sloužil pouze jako mezistruktura pro generování a následné ukládání dat do různých formátů.



Obrázek 8.1-3 Srovnání úložných formátů podle velikosti a rychlosti čtení.

Soubory CSV a HDF5 vykazují (obrázek 8.1-3) vysokou účinnost komprese, která výrazně snižuje jejich velikost při zabalení do ZIP, což může být užitečné zejména ve scénářích vyžadujících optimalizaci ukládání. Naproti tomu soubory XLSX jsou prakticky nekomprimovatelné a jejich velikost v ZIP zůstává srovnatelná s originálem, což je činí méně výhodnými pro použití ve velkých objemech dat nebo v prostředích, kde je důležitá rychlosť přístupu k datům. Kromě toho je doba čtení souborů XLSX ve srovnání s ostatními formáty výrazně vyšší, což je činí méně výhodnými pro rychlé operace čtení dat. Apache Parquet prokázal vysoký výkon pro analytické úlohy a velké objemy dat díky své sloupcové struktuře.

Optimalizace úložiště pomocí aplikace Apache Parquet

Jedním z oblíbených formátů pro ukládání a zpracování velkých dat je Apache Parquet. Tento formát je navržen speciálně pro sloupcové ukládání (podobně jako Pandas), což umožňuje výrazně snížit paměťovou náročnost a zvýšit rychlosť analytických dotazů. Na rozdíl od tradičních formátů, jako jsou CSV a XLSX, podporuje Parquet nativní kompresi a je optimalizován pro systémy pro ukládání velkých objemů dat včetně Sparku, Hadoopu a cloudových úložišť.

Mezi klíčové vlastnosti systému Parquet patří podpora komprese a kódování dat, která výrazně snižuje velikost úložiště a urychluje operace čtení dat tím, že pracuje přímo s požadovanými sloupci, nikoli se všemi řádky dat.

Pro názorný příklad, jak snadno lze získat kód potřebný k převodu dat do formátu Apache Parquet, použijme LLM.

- Pošlete textovou žádost do chatu LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

Napište kód pro uložení dat z Pandas DataFrame do Apache Parquet. ↵

- Odpověď společnosti LLM :

```

1 import pandas as pd
2
3 # Create a DataFrame with construction data
4 construction_data = {
5     'project_id': [1, 2, 3, 4],
6     'project_name': ['Tower A', 'Bridge B', 'Road C', 'Tunnel D'],
7     'location': ['City X', 'City Y', 'City Z', 'City W'],
8     'cost_million_usd': [200, 150, 100, 300]
9 }
10 df_projects = pd.DataFrame(construction_data)
11
12 # Write the DataFrame to the parquet file
13 parquet_path_projects = '/tmp/construction_data/projects.parquet'
14 df_projects.to_parquet(parquet_path_projects)
15
16 print("Construction project data saved in Lakehouse").
17

```

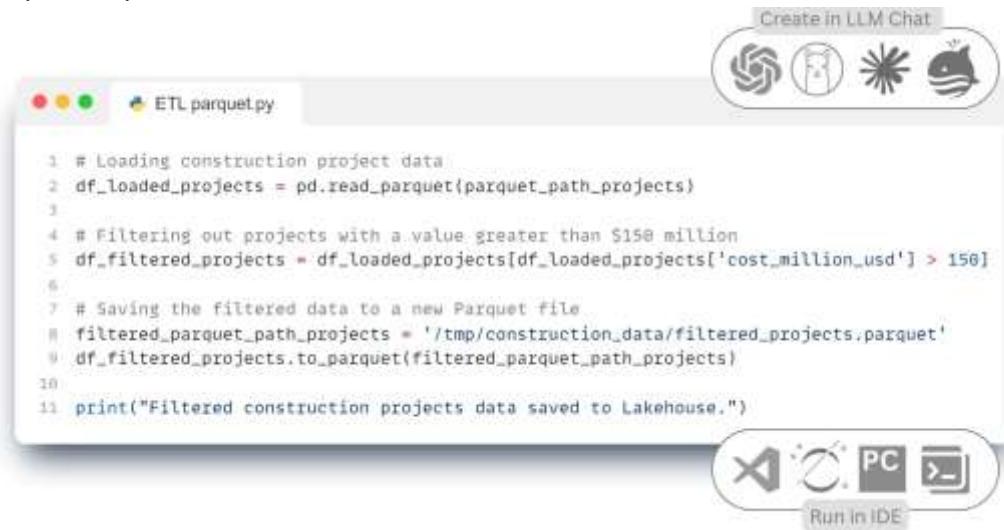
Obrázek 8.1-4 Přenos dat datového rámce z paměti RAM do formátu Apache Parquet pomocí několika řádků jazyka Python .

Další příklad: Simulujme proces ETL s daty uloženými ve formátu Parquet pro filtrování projektů podle určité hodnoty jednoho z atributů "cost_million_usd" (obr. 8.1-4).

- V průběhu chatu pošlete textovou žádost na adresu LLM:

Napište kód, ve kterém chceme filtrovat data v tabulce a uložit pouze ty projekty (řádky tabulky) z dat Apache Parquet, jejichž náklady (parametr cost_million_usd) přesahují 150 milionů USD. ↴

⌚ Odpověď společnosti LLM :



```
# Loading construction project data
df_loaded_projects = pd.read_parquet(parquet_path_projects)

# Filtering out projects with a value greater than $150 million
df_filtered_projects = df_loaded_projects[df_loaded_projects['cost_million_usd'] > 150]

# Saving the filtered data to a new Parquet file
filtered_parquet_path_projects = '/tmp/construction_data/filtered_projects.parquet'
df_filtered_projects.to_parquet(filtered_parquet_path_projects)

print("Filtered construction projects data saved to Lakehouse.")
```

The screenshot shows a code editor window titled "ETL parquet.py". The code itself is a Python script for performing ETL operations on Apache Parquet files. It starts by loading data from a Parquet file, then filters it to keep only projects where the cost is greater than 150 million USD, and finally saves the filtered data back to a new Parquet file. The code editor interface includes a toolbar with icons for "Create in LLM Chat", "Run in IDE", and other development tools.

Obrázek 8.1-5 Proces ETL při práci s daty ve formátu Apache Parquet vypadá stejně jako u jiných strukturovaných formátů.

Použití formátu Parquet (ve srovnání s XLSX, CSV atd.) výrazně snižuje množství uložených informací a urychluje vyhledávání. Díky tomu je vynikající jak pro ukládání, tak pro analýzu dat. Formát Parquet se integruje s různými systémy zpracování a poskytuje efektivní přístup v hybridních architekturách.

Efektivní formát úložiště je však pouze jedním z prvků kompletní datové zkušenosti. K vytvoření udržitelného a škálovatelného prostředí je zapotřebí dobře navržená architektura správy dat. Tuto funkci plní systémy třídy DWH (Data Warehouse). Poskytují agregaci dat z heterogenních zdrojů, transparentnost podnikových procesů a možnost komplexní analýzy pomocí nástrojů BI a algoritmů strojového učení.

DWH: Datový sklad datové sklady

Stejně jako je formát Parquet optimalizován pro efektivní ukládání velkého množství informací, je datový sklad optimalizován pro integraci a strukturování dat pro podporu analytiky, prognózování a manažerského rozhodování.

V dnešních podnicích pocházejí data z mnoha různorodých zdrojů: ERP, CAFM, CPM, CRM systémů, účetnictví a skladového hospodářství, digitálních CAD modelů budov, senzorů IoT a dalších řešení. Pro získání uceleného obrazu nestačí data pouze shromažďovat - je třeba je uspořádat, standardizovat a centralizovat v jednom úložišti. Přesně k tomu slouží DWH - centralizované úložiště, které umožňuje

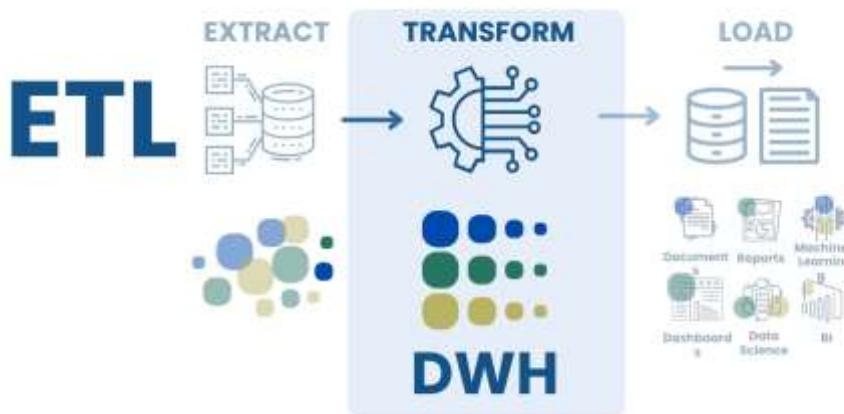
agregovat informace z různých zdrojů, strukturovat je a zpřístupnit pro analytiku a strategické řízení.

DWH (Data Warehouse) je centralizovaný systém datového skladu, který shromažďuje informace z různých zdrojů, strukturuje je a zpřístupňuje pro analytiku a reporting.

V mnoha firmách jsou data rozptýlena v různých systémech, o kterých jsme hovořili v prvních částech knihy (obrázek 1.2-4). DWH tyto zdroje integruje, čímž zajišťuje naprostou přehlednost a spolehlivost informací. Datový sklad DWH je specializovaná databáze (velká databáze), která shromažďuje, zpracovává a uchovává data z různých zdrojů. Hlavními charakteristikami DWH jsou:

- **Použití ETL -procesů** (Extract, Transform, Load) - extrakce dat ze zdrojů, jejich čištění, transformace, načítání do úložiště a automatizace těchto procesů, které byly probrány v sedmé části knihy.
- **Granularita dat** - data v DWH mohou být uložena jak v agregované podobě (souhrnné sestavy), tak v granulární podobě (surová data). Od roku 2024 začínají o granulovaných datech hovořit právě dodavatelé CAD- [125], což možná naznačuje, že se odvětví připravuje na přechod ke specializovaným cloudovým úložištěm pro zpracování dat digitálních modelů budov.
- **Podpora analytiky a prediktivní** - datové sklady poskytují základ pro nástroje BI, Big Data - analýzu a strojové učení.

DWH slouží jako základ pro business intelligence, umožnuje analýzu klíčových ukazatelů výkonnosti, prognózování prodeje, nákupu a nákladů a automatizované reportování a vizualizaci dat (obrázek 8.1-6).



Obr. 8.1-6 V procesu ETL může DWH fungovat jako centrální úložiště, kde data získaná z různých systémů procházejí fázemi transformace a vykládání.

DWH hraje klíčovou roli při integraci, čištění a strukturování informací a poskytuje pevný základ pro business intelligence a rozhodovací procesy. V dnešním prostředí, kde objemy dat rychle rostou a zdroje dat jsou stále rozmanitější, však tradiční přístup DWH k ukládání informací často vyžaduje rozšíření v podobě ELT a Data Lake

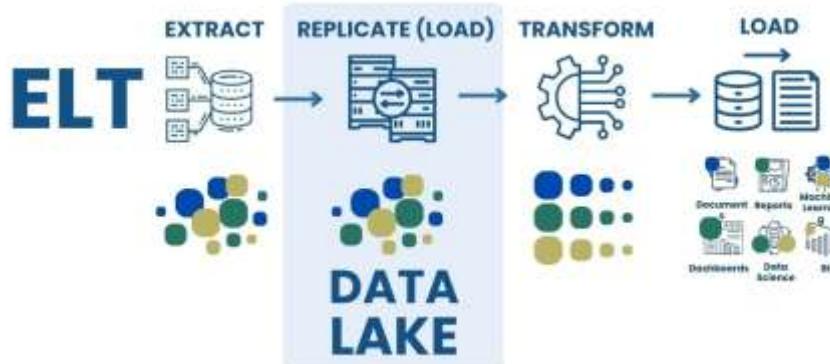
Data Lake - vývoj ETL na ELT: od tradičního čištění k flexibilnímu zpracování.

Klasické DWH - datové sklady, určené k ukládání strukturovaných dat ve formátu optimalizovaném pro analytické dotazy, narážely na omezení při zpracování nestrukturovaných dat a škálovatelnosti. V reakci na tyto problémy vznikla datová jezera), která nabízejí flexibilní úložiště pro velké objemy heterogenních dat.

Data Lake nabízí alternativní přístup k DWH, který umožňuje pracovat s nestrukturovanými, částečně strukturovanými a surovými daty bez předchozího pevného schématu. Tento způsob ukládání je často relevantní pro zpracování dat v reálném čase, strojové učení a pokročilou analýzu. Na rozdíl od DWH, který data před načtením strukturuje a agreguje, Data Lake umožňuje ukládat informace v surové podobě, a tím poskytuje flexibilitu a škálovatelnost

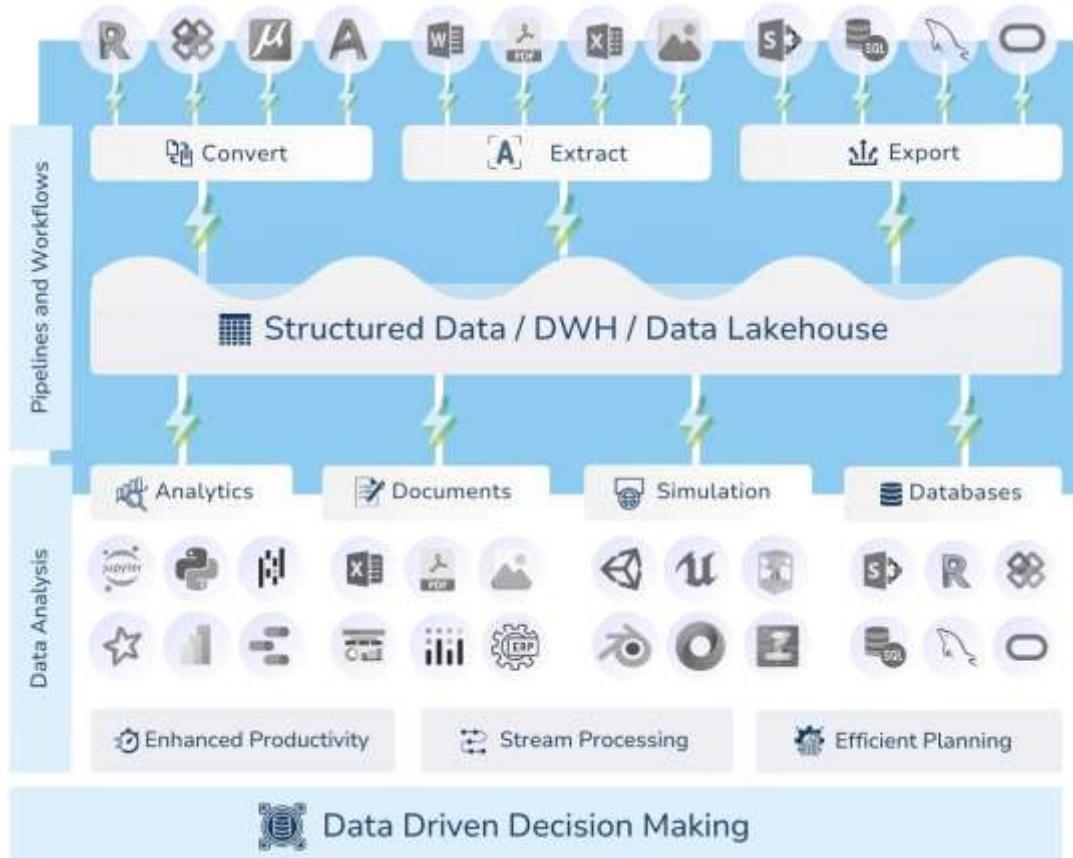
Právě zklamání z tradičních datových skladů (RDBMS, DWH) a zájem o "big data" vedly ke vzniku datových jezer, kde se namísto složitého ETL data nyní jednoduše nahrávají do volně strukturovaného úložiště a zpracování probíhá ve fázi analýzy:

- V tradičních datových skladech se data před vložením do skladu obvykle předem zpracovávají, transformují a čistí (ETL - Extract, Transform, Load) (obrázek 8.1-6). To znamená, že data jsou strukturována a optimalizována pro konkrétní budoucí analytické a reportovací úlohy. Důraz je kladen na zachování vysokého výkonu dotazů a integrity dat. Tento přístup však může být nákladný a méně flexibilní z hlediska integrace nových typů dat a rychle se měnících datových schémat.
- Naproti tomu datová jezera jsou určena k ukládání velkého množství nezpracovaných dat v původním formátu (obrázek 8.1-7). Proces ETL (Extract, Transform, Load), je nahrazován procesem ELT (Extract, Load, Transform), kdy jsou data nejprve načtena do datového skladu "tak, jak jsou", a teprve poté mohou být podle potřeby transformována a analyzována. To poskytuje větší flexibilitu a možnost ukládat heterogenní data, včetně nestrukturovaných dat, jako je text, obrázky a protokoly.



Obr. 8.1-7 Na rozdíl od ETL, Data Lake využívá ELT, v němž jsou informace nejprve nahrány v "surové" podobě a transformace se provádí ve fázi nahrávání.

Tradiční datové skladы se zaměřují na předběžné zpracování dat, aby zajistily vysoký výkon dotazů, zatímco datová jezera upřednostňují flexibilitu: ukládají surová data a transformují je podle potřeby (obrázek 8.1-8).



Obr. 8.1-8 Moderní koncepce ukládání dat mají za cíl ukládat a zpracovávat všechny typy dat pro účely rozhodování.

Navzdory všem výhodám však datová jezera nejsou bez nevýhod. Absence přísné struktury a složitost správy informací může vést k chaosu, v němž se data duplikují, vzájemně si odporuji nebo se stávají irrelevantními. Vyhledávání a analýza dat v takovém úložišti navíc vyžaduje značné úsilí, zejména pokud se jedná o heterogenní informace. K překonání těchto omezení a spojení nejlepších vlastností tradičních datových skladů a datových jezer byla vyvinuta architektura Data Lakehouse.

Architektura datových skladů: synergie skladů a datových jezer

Pro spojení nejlepších vlastností DWH (strukturovaná, spravovatelná, vysoce výkonná analytika) a Data Lake (škálovatelnost, zpracování heterogenních dat) byl vyvinut přístup Data Lakehouse. Tato architektura kombinuje flexibilitu datových jezer s výkonnými nástroji pro zpracování a správu dat typickými pro tradiční datové skladы a dosahuje rovnováhy mezi ukládáním, analýzou a strojovým učením. Data Lakehouse je syntézou datových jezer a datových skladů, která kombinuje flexibilitu a škálovatelnost prvně jmenovaných s možností správy a optimalizace dotazů druhých.

Data Lakehouse je architektonický přístup, který se snaží spojit flexibilitu a škálovatelnost datových jezer se spravovatelností a výkonností dotazů datových skladů (obrázek 8.1-9).

Mezi klíčové vlastnosti Data Lakehouse patří:

- **Otevřený formát pro ukládání dat:** používání otevřených formátů pro ukládání dat, jako je například Apache Parquet, zajišťuje efektivitu a optimalizaci dotazů.
- **Schéma pouze pro čtení:** Na rozdíl od tradičního přístupu schématu pouze pro zápis v DWH, Lakehouse podporuje schéma pouze pro čtení, které umožňuje větší flexibilitu při správě datové struktury.
- **Flexibilní a škálovatelný:** podporuje ukládání a analýzu strukturovaných i nestrukturovaných dat a poskytuje vysoký výkon dotazů díky optimalizaci na úrovni úložiště.

Data Lakehouse nabízí kompromisní řešení, které kombinuje výhody obou přístupů, takže je ideální pro moderní analytické úlohy, které vyžadují flexibilitu při zpracování dat.



Obr. 8.1-9 Data Lakehouse je nová generace úložných systémů navržených tak, aby splňovaly komplexní a neustále se měnící požadavky.

Myšlenka moderních datových skladů se zdá být jednoduchá: pokud jsou všechna data na jednom místě, je snazší je analyzovat. V praxi však není vše tak hladké. Představte si, že se firma rozhodne zcela opustit běžné účetní a manažerské systémy (ERP, PMIS, CAFM nebo jiné) a nahradí je jedním obrovským datovým jezerem, ke kterému mají všichni přístup. Co se stane? S největší pravděpodobností nastane chaos: data budou duplicitní, rozporuplná a kritické informace se ztratí nebo poškodí. I kdyby se datové jezero používalo pouze pro analytické účely, bez řádné správy bude vážně ohroženo:

- Data jsou obtížně srozumitelná: v běžných systémech mají data jasnou strukturu, ale v jezeře jsou jen obrovskou kumulací souborů a tabulek. Aby něco našel, musí zjistit, za co jsou

jednotlivé řádky a sloupce zodpovědné.

- Data mohou být nepřesná: pokud je na jednom místě uloženo mnoho verzí stejných informací, je obtížné zjistit, která verze je aktuální. V důsledku toho jsou rozhodnutí přijímána na základě zastaralých nebo chybných údajů.
- Příprava dat pro práci je obtížná: data je třeba nejen uložit, ale také prezentovat ve vhodné formě - v podobě sestav, grafů, tabulek. V tradičních systémech se to provádí automaticky, ale v datových jezerech to vyžaduje další zpracování.

Každý koncept datového skladu má proto své vlastní charakteristiky, přístupy ke zpracování a obchodní aplikace. Tradiční databáze se zaměřují na transakční operace, datové sklady (DWH) poskytují strukturu pro analytiku, datová jezera (Data Lake) ukládají informace v surové podobě a hybridní sklady (Data Lakehouse) kombinují výhody DWH a Data Lake (obrázek 8.1-10).

	Traditional Approach	Data Warehouse	Data Lake	Data Lakehouse
Data Types	Relational Databases	Structured, ready for analytics	Raw, semi-structured, or unstructured	Mix of structured and unstructured
Use Cases	Transactional Systems	Reporting, dashboards, BI	Big data storage, AI, advanced analytics	Hybrid analytics, AI, real-time data
Processing	OLTP – real-time transactions	ETL – clean and structure before analysis	ELT – store raw data, transform later	ELT with optimized storage and real-time processing
Storage	On-premise servers	Centralized, SQL-based	Decentralized, flexible formats	Combines advantages of DWH and DL
Common Tools	MySQL, PostgreSQL	Snowflake, Redshift, BigQuery	Hadoop, AWS S3, Azure Data Lake	Databricks, Snowflake, Google BigLake

Obr. 8.1-10 DWH, Data Lake a Data Lakehouse: hlavní rozdíly v typech dat, scénářích použití, metodách zpracování a přistupech k ukládání.

Výběr architektury úložiště je složitý proces, který závisí na potřebách podniku, objemu informací a požadavcích na analýzu. Každé řešení má své výhody a nevýhody: DWH poskytuje strukturu, Data Lake poskytuje flexibilitu a Lakehouse poskytuje rovnováhu mezi oběma. Organizace jsou jen zřídka omezeny na jedinou datovou architekturu.

Bez ohledu na zvolenou architekturu jsou automatizované systémy správy dat výrazně lepší než manuální metody. Minimalizují lidské chyby, urychlují zpracování informací a zajišťují transparentnost a sledovatelnost dat ve všech fázích podnikových procesů.

Zatímco centralizované datové skladы se již staly průmyslovým standardem v mnoha oblastech hospodářství, situace ve stavebnictví zůstává roztříštěná. Data zde jsou distribuována na různých platformách (CDE, PMIS, ERP atd.), což ztěžuje vytvoření jednotného obrazu o dění a vyžaduje architektury schopné tyto zdroje spojit do uceleného, analyticky využitelného digitálního prostředí.

CDE, PMIS, ERP nebo DWH a datové jezero.

Některé stavební a inženýrské společnosti již používají koncept společného datového prostředí (CDE) podle normy ISO 19650. CDE plní v podstatě stejné funkce jako datový sklad (DWH) v jiných odvětvích: centralizuje informace, zajišťuje kontrolu verzí, poskytuje přístup k ověřeným informacím.

Společné datové prostředí (CDE) je centralizovaný digitální prostor, který slouží ke správě, ukládání, sdílení a spolupráci na projektových informacích ve všech fázích životního cyklu objektu. CDE je často implementováno pomocí clouдовých technologií a integrováno se systémy CAD (BIM).

Finanční, maloobchodní, logistické a průmyslové odvětví již desítky let využívají centralizované systémy správy dat, které kombinují informace z různých zdrojů, kontrolují jejich relevanci a poskytují analytiku. CDE posouvá tyto principy dále tím, že je přizpůsobuje výzvám navrhování a řízení životního cyklu budov.

Stejně jako DWH, CDE strukturuje data, zachycuje změny a poskytuje jednotné místo přístupu k ověřeným informacím. S přechodem na cloud a integrací s analytickými nástroji jsou rozdíly mezi nimi stále méně patrné. Přidáme-li k CDE granulární data, o jejichž konceptu se v CAD -prodejnách hovoří již od roku 2023[93, 125], můžeme vidět ještě více paralel s klasickým DWH.

Již dříve jsme se v kapitole "Stavební ERP a PMIS systémy" zabývali systémy PMIS (Project Management Information System) a ERP (Enterprise Resource Planning). Ve stavebních projektech CDE a PMIS spolupracují: CDE slouží jako úložiště dat včetně výkresů, modelů a projektové dokumentace, zatímco PMIS řídí procesy, jako je řízení termínů, úkolů, zdrojů a rozpočtů.

Systém ERP, který odpovídá za řízení podniku jako celku (finance, nákup, personalistika, výroba), lze integrovat se systémem PMIS, který zajišťuje kontrolu nákladů a rozpočtu na úrovni podniku. Pro analytiku a reporting lze DWH použít ke shromažďování, strukturování a agregaci dat z CDE, PMIS a ERP k vyhodnocování finančních klíčových ukazatelů výkonnosti (ROI) a identifikaci vzorců. Datové jezero (DL) zase může doplnit DWH ukládáním nezpracovaných a nestrukturovaných dat (např. protokolů, dat ze senzorů, obrázků). Tato data lze zpracovat a načíst do DWH pro další analýzu.

CDE a PMIS se tedy zaměřují na řízení projektů, ERP na podnikové procesy a DWH a Data Lake na analytiku a datové operace.

Při porovnání systémů CDE, PMIS a ERP se systémy DWH a Data Lake jsou patrné významné rozdíly z hlediska nezávislosti na dodavateli, nákladů, flexibility integrace, nezávislosti na datech, rychlosti přizpůsobení se změnám a analytických schopností (obrázek 8.1-11). Tradiční systémy, jako jsou CDE, PMIS a ERP, jsou často vázány na konkrétní dodavatelská řešení a standardy, což je činí méně flexibilními a zvyšuje jejich náklady kvůli licencím a podpoře. Data v těchto systémech jsou navíc často zapouzdřena v proprietárních, uzavřených formátech, což omezuje jejich využití a analýzu.

		CDE, PMIS, ERP	DWH, Data Lake
	Vendor Dependency	High (tied to specific solutions and standards of vendors)	Low (flexibility in tool and platform choice)
	Integration Flexibility	Limited (integration depends on vendor solutions)	High (easily integrates with various data sources)
	Cost	High (licensing and support costs)	Relatively lower (use of open technologies and platforms)
	Data Independence	Low (data often locked in proprietary formats)	High (data stored in open and accessible formats)
	Adaptability to Changes	Slow (changes require vendor approval and integration)	Fast (adaptation and data structure modification without intermediaries)
	Analytical Capabilities	Limited (dependent on vendor-provided solutions)	Extensive (support for a wide range of analytical tools)

Obr. 8.1-11 DWH a Data Lake nabízejí větší flexibilitu a datovou nezávislost než systémy jako CDE, PMIS a ERP.

Naproti tomu DWH a Data Lake poskytují větší flexibilitu při integraci s různými zdroji dat a jejich využití otevřených technologií a platem pomáhá snížit celkové náklady na vlastnictví. DWH a Data Lake navíc podporují širokou škálu analytických nástrojů, což zvyšuje možnosti analýzy a správy.

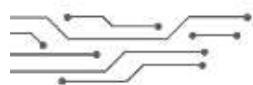
S rozvojem nástrojů pro reverzní inženýrství pro formáty CAD a přístupem k databázím aplikací CAD se stále naléhavěji klade otázka: nakolik je opodstatněné pokračovat v používání uzavřených, izolovaných platem, pokud musí být konstrukční data k dispozici širokému spektru specialistů pracujících v desítkách dodavatelských a projekčních organizací?

Tato technologická závislost na konkrétním dodavateli může výrazně omezit flexibilitu správy dat, zpomalit reakce na změny v projektu a bránit efektivní spolupráci mezi účastníky.

Tradiční přístupy ke správě dat - včetně DWH, Data Lake, CDE a PMIS - se zaměřují především na ukládání, strukturování a zpracování informací. S rozvojem umělé inteligence a strojového učení však roste potřeba nových způsobů organizace dat, které nejen agregují, ale také odhalují složité vztahy, nacházejí skryté vzorce a poskytují okamžitý přístup k nejdůležitějším informacím.

Zvláštní roli v tomto směru začínají hrát vektorové databáze - nový typ úložiště optimalizovaný pro

vysokorozměrná vnoření.



KAPITOLA 8.2.

SPRÁVA DATOVÉHO SKLADU A PREVENCE CHAOSU

Vektorové databáze a ohraničující pole

Vektorové databáze jsou novou třídou úložišť, která neukládají pouze data, ale umožňují vyhledávání podle významu, porovnávání objektů podle sémantické blízkosti a vytváření inteligentních systémů: od doporučení po automatickou analýzu a generování kontextu. Na rozdíl od tradičních databází, které se zaměřují na přesné shody, vektorové databáze vyhledávají podobné objekty na základě atributů - i když neexistuje přesná shoda

Vektorová databáze je specializovaný typ databáze, která ukládá data jako vícerozměrné vektory, z nichž každý představuje určité charakteristiky nebo vlastnosti. Tyto vektory mohou mít různý počet dimenzí v závislosti na složitosti dat (v jednom případě to může být několik dimenzí a v jiném — tisíce).

Hlavní výhodou vektorových databází je vyhledávání podle sémantické relevance, nikoli podle přesné shody hodnot. Místo SQL- a Pandas-dotazů s filtry "rovná se" nebo "obsahuje" se používá vyhledávání nejbližších sousedů (k-NN) (o k-NN si povíme více v další části knihy) v prostoru příznaků.

S rozvojem LLM (Large Language Models) a generativních modelů se interakce s databázemi začíná měnit. Nyní je možné se dotazovat na data v přirozeném jazyce, získat sémantické vyhledávání v dokumentech, automaticky extrahovat klíčové výrazy a vytvářet kontextové vztahy mezi objekty - to vše bez nutnosti ovládat jazyk SQL nebo znát strukturu tabulek. Podrobněji o tom bylo pojednáno v části "LLM a jejich role při zpracování dat a v obchodních procesech".

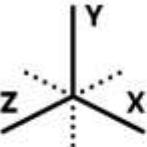
Je však důležité si uvědomit, že LLM automaticky nestrukturuje a neuspořádává informace. Model pouze proplouvá daty a na základě kontextu dotazu vyhledává nejrelevantnější část dat. Pokud data nebyla předem vyčištěna nebo transformována, bude hlubkové vyhledávání podobné snaze najít odpověď v digitálním "smetí" - může fungovat, ale kvalita výsledků bude nižší. Ideální je, pokud lze data strukturovat (např. přeložit dokumenty do formátu Markdown) a načíst je do vektorové databáze. Tím se výrazně zvýší přesnost a relevance výstupu.

Původně se vektorové databáze používaly ve strojovém učení, ale dnes nacházejí stále více aplikací i mimo něj - ve vyhledávačích, personalizaci obsahu a intelligentní analytice.

Jedním z nejzřetelnějších příkladů vektorového přístupu v konstrukci je ohraničující box (bounding parallelepiped). Jedná se o geometrickou konstrukci popisující hranice objektu v trojrozměrném prostoru. Bounding Box je definován minimálními a maximálními souřadnicemi X, Y a Z, které tvoří "krabici" kolem objektu. Tato metoda umožňuje odhadnout velikost a umístění prvku, aniž by bylo nutné analyzovat celou geometrii.

Každý ohraničující rámeček může být reprezentován jako vektor ve vícerozměrném prostoru: například

[x, y, z, šířka, výška, hloubka] - již 6 rozměrů (obr. 8.2-1).



Bounding Box

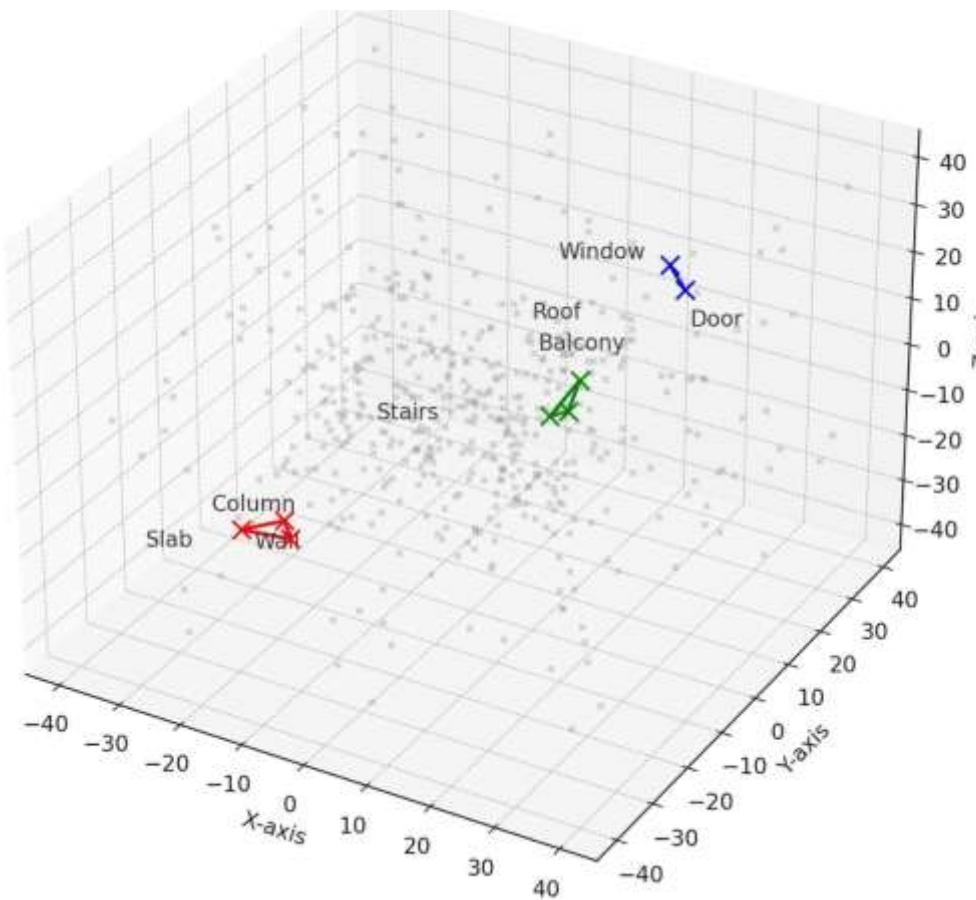
	minX	maxX	minY	maxY	minZ	maxZ	Width	Height	Depth
Column	-15	-5	-25	-15	0	10	10	10	20
Stairs	-5	5	-15	-5	0	10	10	10	10
Door	5	15	5	15	0	10	10	10	10
Window	25	35	-35	-25	10	30	10	20	20
Balcony	15	25	-5	5	20	40	10	20	20

Obr. 8.2-1 Bounding Box -informace o souřadnicích prvků a jejich umístění v modelu projektu je analogická vektorové databázi.

Tato reprezentace dat usnadňuje mnoho úkolů, včetně kontroly průsečíků mezi objekty, plánování prostorového rozmístění stavebních prvků a provádění automatických výpočtů. Bounding Box může sloužit jako most mezi složitými 3D modely a tradičními vektorovými databázemi, což umožnuje efektivně využívat výhody obou přístupů při architektonickém a inženýrském modelování

Bounding Box je "vektorizace geometrie" a embedding (způsob transformace něčeho abstraktního) je "vektorizace významu". Oba přístupy umožňují přejít od ručního vyhledávání k inteligentnímu vyhledávání, ať už se jedná o 3D objekty v modelu projektu nebo pojmy v textu.

Vyhledávání objektů v projektu (například "najdi všechna okna o šířce > 1,5 m") je podobné vyhledávání nejbližších sousedů (k-NN) ve vektorové databázi, kde kritéria definují "zónu" v prostoru příznaků. (o vyhledávání nejbližších sousedů k-NN budeme více hovořit v další části o strojovém učení) (obr. 8.2-2). Přidáme-li k atributům ohrazení další parametry (materiál, hmotnost, doba výroby), změní se tabulka na vysokorozměrný vektor, kde každý atribut představuje novou dimenzi. To se blíží moderním vektorovým bázím, kde se dimenze počítají na stovky nebo tisíce (např. vložení z neuronových sítí).



Obr. 8.2-2 Vyhledávání objektů v projektu pomocí vektorových databází.

Přístup použitý v Bounding Box, je použitelný nejen pro geometrické objekty, ale také pro analýzu textu a jazyka. Vektorové reprezentace dat se již aktivně používají při zpracování přirozeného jazyka (NLP). Stejně jako lze objekty ve stavebním projektu seskupit podle jejich prostorové blízkosti (obrázek 8.2-2), lze slova v textu analyzovat podle jejich sémantické a kontextové blízkosti.

Například slova "architekt", "konstrukce", "design" budou ve vektorovém prostoru vedle sebe, protože mají podobný význam. V LLM tento mechanismus umožňuje automatickou, nevyžadující ruční kategorizaci:

- Určit téma textu
- Provádění sémantického vyhledávání v obsahu dokumentů
- Generování automatických anotací a textových shrnutí
- Vyhledání synonym a souvisejících termínů

Vektorové databáze umožňují analyzovat text a vyhledávat v něm související pojmy stejným způsobem, jakým vám Bounding Box pomáhá analyzovat prostorové objekty ve 3D -modelu. Příklad Bounding Box prvků projektu pomáhá pochopit, že vektorová reprezentace není čistě "umělý" koncept z ML, ale přirozený způsob strukturování dat pro řešení aplikačních problémů, ať už jde o vyhledávání

sloupců v CAD projektu nebo sémanticky blízkých obrázků v databázi.

Odborníci pracující s databázemi by měli věnovat pozornost vektorovým úložištím. Jejich rozšíření naznačuje novou etapu vývoje databází, kdy se začínají prolínat klasické relační systémy a technologie orientované na umělou inteligenci a vytvářejí hybridní řešení budoucnosti.

Uživatelé, kteří vyvíjejí složité a rozsáhlé aplikace umělé inteligence, budou používat specializované databáze pro vektorové vyhledávání. Současně ti, kteří potřebují pouze samostatné AI-funkce pro integraci do stávajících aplikací, budou spíše volit vestavěné možnosti vektorového vyhledávání v databázích, které již používají (PostgreSQL, Redis).

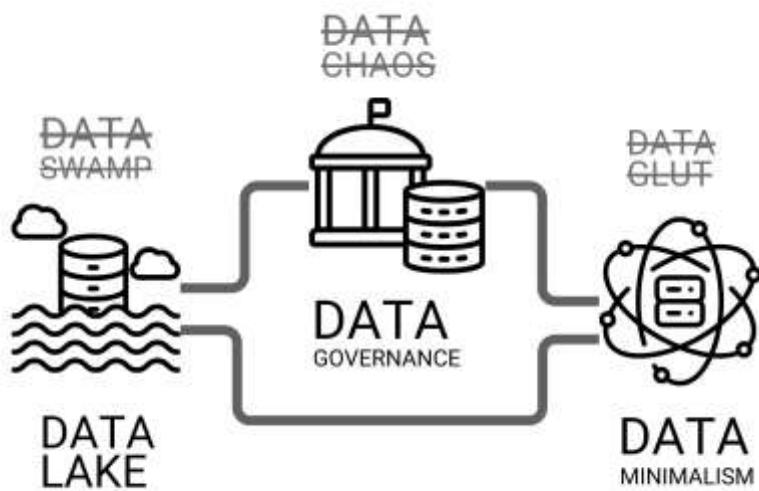
Přestože systémy jako DWH, Data Lake, CDE, PMIS, vektorové databáze a další nabízejí různé přístupy k ukládání a správě dat, jejich účinnost je dána nejen jejich architekturou, ale také tím, jak dobře jsou samotná data organizována a spravována. I při použití moderních řešení - ať už jde o vektorové databáze, klasické relační DBMS nebo sklady typu Data Lake - může absence jasných pravidel pro správu, strukturování a aktualizaci dat vést ke stejným potížím, s jakými se potýkají uživatelé pracující s různorodými soubory a daty různých formátů.

Bez správy dat) se i ta nejvýkonnější řešení mohou stát chaotickými a nestrukturovanými, což z datových jezer udělá datové bažiny). Aby se tomu společnosti vyhnuly, musí nejen zvolit správnou architekturu úložiště, ale také zavést strategii minimalizace dat), řízení přístupu a kontroly kvality, aby se data stala účinným nástrojem pro rozhodování.

Data Governance), Data Minimalism) a Data Swamp).

Pochopení a zavedení koncepcí správy dat), minimalizace dat) a prevence datové bažiny) jsou klíčem k úspěšnému řízení datových skladů a poskytování obchodní hodnoty (obrázek 8.2-3).

Podle studie společnosti Gartner (2017) selže 85% projektů big data a jedním z klíčových důvodů je nedostatečná kvalita dat a jejich správa [144].



Obr. 8.2-3 Mezi klíčové aspekty správy dat patří správa dat a minimalizace dat.

Správa dat (Data Governance) je základní složkou správy dat, která zajišťuje, že data jsou ve všech podnikových procesech využívána vhodně a efektivně. Nejde jen o stanovení pravidel a postupů, ale také o zajištění dostupnosti, spolehlivosti a bezpečnosti dat:

- Definice a klasifikace dat: Jasná definice a klasifikace entit umožňuje organizacím pochopit, jaké entity jsou ve společnosti potřebné, a určit, jak by měly být použity.
- Přístupová práva a správa: vytvoření zásad a postupů pro přístup k údajům a jejich správu zajišťuje, že k určitým údajům mají přístup pouze oprávnění uživatelé.
- Ochrana dat před vnějšími hrozbami: Ochrana dat před vnějšími hrozbami je klíčovým aspektem správy dat. To zahrnuje nejen technická opatření, ale také školení zaměstnanců v základech bezpečnosti informací.

Minimalismus dat (Data Minimalism) je přístup k redukci dat na nejcennější a nejvýznamnější atributy a entity v sestavě (obrázek 8.2-4), címž se snižují náklady a zlepšuje využití dat:

- Zjednodušení rozhodování: redukce počtu objektů a jejich atributů na ty nejdůležitější zjednodušuje rozhodování tím, že snižuje čas a zdroje potřebné k analýze a zpracování dat.
- Zaměření na to, co je důležité: výběr nejdůležitějších entit a atributů umožňuje zaměřit se na informace, které jsou pro firmu skutečně důležité, a eliminovat šum a nepotřebná data.
- Efektivní přidělování zdrojů: minimalizace dat umožňuje efektivnější přidělování zdrojů, snižuje náklady na ukládání a zpracování dat, zlepšuje kvalitu a bezpečnost dat.

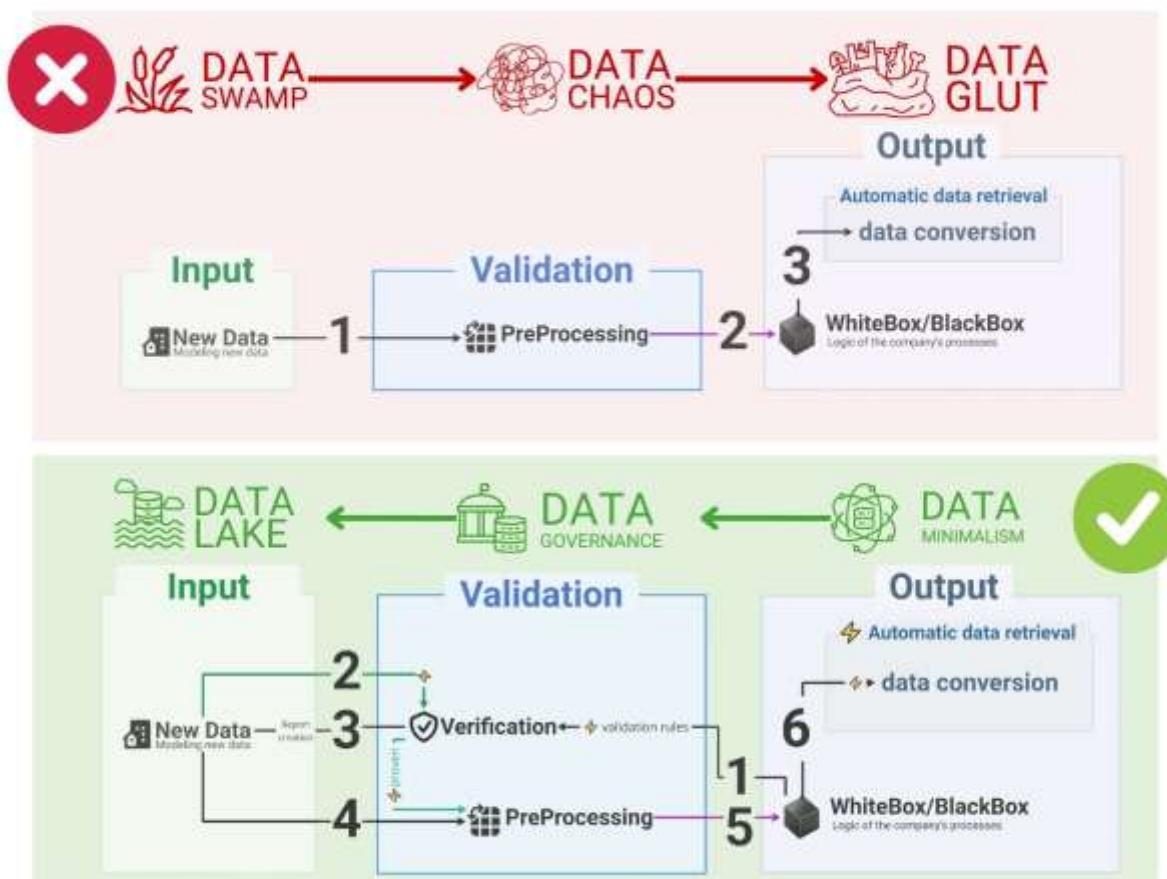
Logika práce s daty by neměla začínat jejich samotným vytvořením (obr. 8.2-4), ale pochopením budoucích scénářů využití těchto dat ještě před zahájením procesu jejich generování. Tento přístup umožňuje předem definovat minimální nezbytné požadavky na atributy, jejich typy a hraniční hodnoty. Tyto požadavky tvoří základ pro vytvoření správných a stabilních entit v informačním modelu. Předběžné pochopení účelů a použití dat přispívá k vytvoření struktury vhodné pro analýzu. Podrobněji o přístupech k modelování dat na konceptuální, logické a fyzické úrovni bylo pojednáno v kapitole "Modelování dat: konceptuální, logický a fyzický model".

V tradičních podnikových procesech stavebních firem zpracování dat častěji připomíná vysypání dat do bažiny, kdy se data nejprve vytvoří a poté se je specialisté snaží integrovat do dalších systémů a nástrojů.

Datová bažina (Data Swamp) je výsledkem nekontrolovaného shromažďování a ukládání dat bez řádné organizace, strukturování a správy, což vede k tomu, že data jsou nestrukturovaná, obtížně využitelná a málo hodnotná.

Jak zabránit tomu, aby se tok informací změnil v bažinu:

- **Správa struktury dat:** zajištění strukturovanosti a kategorizace dat na pomáhá předcházet zahlcení daty tím, že jsou přehledná a snadno dostupná.
- **Porozumění a interpretace dat:** jasný popis původu, úprav a významu dat zajišťuje, že jsou data správně pochopena a interpretována.
- **Udržování kvality dat:** pravidelná údržba a čištění dat pomáhá udržovat jejich kvalitu, relevanci a hodnotu pro analytické a obchodní procesy.



Obrázek 8.2-4 Abyste se vyhnuli nepořádku v datovém skladu, měli byste proces vytváření dat zahájit shromázděním požadavků na atributy.

Začleněním zásad správy dat a datového minimalismu do procesů správy dat a aktivní prevencí toho, aby se datové sklady staly datovými bažinami, mohou organizace maximalizovat potenciál svých dat.

Další etapou ve vývoji práce s daty je po vyřešení otázek správy a minimalizace standardizace automatického zpracování, zajištění kvality a zavedení metod, které umožňují využití dat pro analýzu, transformaci a rozhodování. Tím se zabývají metodiky DataOps a VectorOps, které se stávají důležitými nástroji pro firmy pracující s velkými daty a strojovým učením.

DataOps a VectorOps: nové datové standardy.

Zatímco Data Governance je zodpovědná za kontrolu a organizaci dat, DataOps pomáhá zajistit jejich přesnost, konzistenci a hladký tok v rámci společnosti. To je obzvláště důležité pro řadu obchodních případů ve stavebnictví, kde data vznikají průběžně a vyžadují včasné zpracování. Například v situacích, kdy je třeba synchronizovat informační modely staveb, požadavky na projekt a analytické zprávy mezi různými systémy během jediného pracovního dne, může být role DataOps klíčová. Umožňuje vybudovat stabilní a opakovatelné procesy zpracování dat, čímž snižuje riziko zpoždění a ztráty relevance informací.

Samotná správa dat nestačí - je nezbytné, aby data nebyla pouze uložena, ale aby byla aktivně využívána v každodenním provozu. Zde přichází ke slovu DataOps - metodika zaměřená na automatizaci, integraci a nepřetržitý tok dat.

DataOps se zaměřuje na zlepšení spolupráce, integrace a automatizace datových toků v organizacích. Přijetí postupů DataOps podporuje přesnost, konzistenci a dostupnost dat, což je pro datově orientované aplikace zásadní.

Klíčovými nástroji v ekosystému DataOps jsou Apache Airflow (obrázek 7.4-4) pro orchestraci pracovních postupů a Apache NiFi (obrázek 7.4-5) pro směrování a transformaci datových toků. Společně tyto technologie umožňují flexibilní, spolehlivé a škálovatelné datové kanály pro automatické zpracování, řízení a integraci informací mezi systémy (více informací v kapitole "Automatický ETL - převodník"). Při implementaci přístupu DataOps do stavebních procesů je důležité zohlednit čtyři základní aspekty:

1. **Lidé a nástroje jsou důležitější než data:** oddělená datová úložiště mohou být považována za hlavní problém, ale skutečnost je složitější. Kromě rozdílnosti dat hraje významnou roli izolace týmů a rozdílné nástroje, které používají. Ve stavebnictví pracují s daty specialisté z různých oborů: datoví inženýři a analytici, týmy BI a vizualizace, stejně jako odborníci na řízení projektů a kvalitu. Každý z nich má odlišné způsoby práce, a proto se stává důležitým vytvořit ekosystém, v němž data mezi účastníky volně proudí a který poskytuje jedinou konzistentní verzi informací.
2. **Automatizace testování a odhalování chyb:** Stavební data vždy obsahují chyby, ať už jde o nepřesnosti v modelech, chyby ve výpočtech nebo zastaralé specifikace. Pravidelné testování

dat a odstraňování opakujících se chyb může výrazně zlepšit kvalitu dat. V rámci DataOps je třeba zavést automatizované kontrolní a validační mechanismy, které monitorují správnost dat, analyzují chyby a identifikují vzory a zachycují a řeší systémové chyby v každém pracovním postupu. Čím vyšší je stupeň automatizované validace, tím vyšší je celková kvalita dat a tím nižší je pravděpodobnost výskytu chyb v konečných fázích.

3. **Data by měla být testována stejně jako programový kód:** většina stavebních aplikací je založena na zpracování dat, ale jejich kontrola je často ponechána na vedlejších rolích. Pokud jsou modely strojového učení trénovány na nepřesných datech, vede to k nesprávným předpovědím a finančním ztrátám. V rámci DataOps by data měla být podrobena stejně přísné kontrole jako programový kód: logické kontroly, zátěžové testy, vyhodnocení chování modelů při změně vstupních hodnot. Pouze validovaná a spolehlivá data mohou být použita jako základ pro manažerská rozhodnutí.
4. **Sledovatelnost dat bez omezení výkonnosti:** sledování dat není jen sbírkou metrik, ale strategickým nástrojem řízení kvality. Aby DataOps fungoval efektivně, musí být pozorovatelnost zabudována ve všech fázích práce s daty, od návrhu až po provoz. Zároveň je důležité, aby monitorování nezpomalovalo systém. Ve stavebních projektech je rozhodující nejen shromažďovat data, ale činit tak takovým způsobem, aby nebyla nijak narušena práce odborníků (např. projektantů), kteří data vytvářejí. Tato rovnováha umožňuje kontrolovat kvalitu dat, aniž by byla ohrožena produktivita.

DataOps není pro datové vědce další zátěží, ale páteří jejich práce. Zavedením DataOps mohou stavební firmy přejít od chaotické správy dat k efektivnímu ekosystému, kde data pracují pro byznys.

VectorOps zase představuje další etapu vývoje DataOps, zaměřenou na zpracování, ukládání a analýzu vícerozměrných vektorových dat (o kterých byla řeč v předchozí kapitole). To má význam zejména v oblastech, jako jsou digitální dvojčata, modely neuronových sítí a sémantické vyhledávání, které začínají přicházet do stavebnictví. VectorOps se opírá o vektorové databáze, které umožňují efektivně ukládat, indexovat a prohledávat vícerozměrné reprezentace objektů.

VectorOps je dalším krokem po DataOps, zaměřeným na zpracování, analýzu a využití vektorových dat ve stavebnictví. Na rozdíl od DataOps, který se zaměřuje na tok, konzistenci a kvalitu dat, se VectorOps zaměřuje na správu vícerozměrných reprezentací objektů potřebných pro strojové učení.

Na rozdíl od tradičních přístupů umožňuje VectorOps dosáhnout přesnějšího popisu objektů, což je důležité pro digitální dvojčata, systémy generativního navrhování a automatickou detekci chyb v datech CAD převedených do vektorového formátu. Kombinovaná implementace DataOps a VectorOps tvoří pevný základ pro škálovatelnou, automatizovanou práci s velkými objemy informací - od klasických tabulek až po sémanticky bohaté prostorové modely.

Další kroky: od chaotického úložiště ke strukturovanému úložišti

Tradiční přístupy k budování datových úložišť často vedou k vytváření různorodých "informačních sil", kde jsou důležité poznatky nedostupné pro analýzu a rozhodování. Moderní koncepty ukládání dat,

jako jsou datové sklady, datová jezera a jejich hybridy, umožňují sjednotit různorodé informace a centralizovaně je zpřístupnit pro datové toky a business intelligence. Důležité je nejen zvolit správnou architekturu úložiště, ale také zavést správu dat) a datový minimalismus), aby se úložiště nestala nekontrolovatelnými datovými bažinami).

Pro shrnutí této části je vhodné zdůraznit hlavní praktické kroky, které vám pomohou aplikovat probrané koncepty na vaše každodenní úkoly:

■ Výběr efektivních formátů pro ukládání dat

- Přechod od formátů CSV a XLSX k efektivnějším formátům (Apache Parquet, ORC) pro ukládání velkého množství dat.
- Zavedení systému verzování dat pro sledování změn
- Používání metadat k popisu struktury a původu informací.

■ Vytvoření jednotné architektury podnikových dat

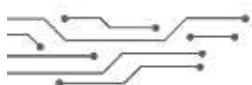
- Porovnání různých architektur úložišť: RDBMS, DWH a Data Lake. Vyberte si tu, která nejlépe vyhovuje vašim potřebám v oblasti škálovatelnosti, integrace zdrojů a analytického zpracování.
- Navrhněte mapu procesů pro extrakci, načítání a transformaci dat (ETL) z různých zdrojů pro své úlohy. K vizualizaci klíčových kroků a integračních bodů použijte vizualizační nástroje, jako je Miro, Lucidchart nebo Draw.io.

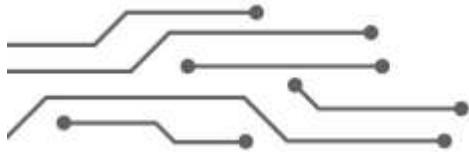
■ Zavedení postupů správy dat a minimalizace dat

- Dodržujte přístup minimalizace dat - ukládejte a zpracovávejte pouze to, co je skutečně cenné.
- Zavedení zásad správy dat - definování odpovědnosti za data, zajištění kvality a transparentnosti.
- Další informace o zásadách správy dat a koncepcích DataOps, VectorOps.
- Definovat kritéria kvality dat a postupy pro validaci dat v rámci DataOps.

Dobře organizované ukládání dat vytváří základ pro centralizaci analytických procesů společnosti. Přechod od chaotického hromadění souborů ke strukturovaným úložištěm umožňuje proměnit informace ve strategické aktivum, které pomáhá přijímat informovaná rozhodnutí a zlepšovat efektivitu podnikových procesů.

Po automatizaci a standardizaci procesů sběru, transformace, analýzy a strukturovaného ukládání dat přichází na řadu další fáze digitální transformace, kterou je plnohodnotné zpracování velkých objemů dat.





IX ČÁST

VELKÁ DATA, STROJOVÉ UČENÍ A PŘEDPOVĚDI

Devátý díl se zaměřuje na big data, strojové učení a prediktivní analýzu ve stavebnictví. Zkoumá přechod od intuitivního rozhodování k objektivním analýzám založeným na historických datech. Na praktických příkladech je demonstrována analýza velkých dat ve stavebnictví - od rozboru datové sady stavebních povolení v San Francisku až po zpracování CAD - projektů s miliony prvků. Zvláštní pozornost je věnována metodám strojového učení pro predikci nákladů a harmonogramu stavebních projektů, přičemž se podrobně rozebírají algoritmy lineární regrese a k-nejbližších sousedů. Je ukázáno, jak se strukturovaná data stávají základem prediktivních modelů pro hodnocení rizik, optimalizaci zdrojů a zlepšení efektivity řízení projektů. V této části jsou také uvedena doporučení pro výběr reprezentativních vzorků dat a vysvětleno, proč pro efektivní analýzy nejsou vždy nutné velké soubory dat.

KAPITOLA 9.1.

VELKÁ DATA A JEJICH ANALÝZA

Velká data ve stavebnictví: od intuice k předvídatelnosti

Pojem "big data" nemá striktní definici. Původně se tento pojem objevil, když objem informací začal přesahovat možnosti tradičních metod jejich zpracování. Dnes se objem a složitost dat v mnoha odvětvích, včetně stavebnictví, zvýšily natolik, že se nevezdou do lokální paměti počítačů a vyžadují použití nových technologií pro jejich zpracování.

Podstatou práce s velkými daty je nejen jejich ukládání a zpracování, ale také prediktivní schopnosti. Ve stavebnictví otevírají Big Data cestu od intuitivních rozhodnutí založených na subjektivní interpretaci tabulek a vizualizací (jak bylo uvedeno výše) k informovaným prognózám podloženým reálnými pozorováními a statistikami.

Navzdory všeobecnému přesvědčení není cílem práce s velkými objemy dat "přimět stroj, aby myslel jako člověk", ale použít matematické modely a algoritmy k analýze obrovského množství dat za účelem identifikace vzorců, předvídání událostí a optimalizace procesů.

Big Data nejsou chladným světem algoritmů bez lidského vlivu. Naopak, big data fungují ve spojení s našimi instinkty, chybami a kreativitou. Právě nedokonalost lidského myšlení nám umožňuje nacházet nestandardní řešení a dělat průlomy.

S rozvojem digitálních technologií začal stavební průmysl aktivně využívat techniky zpracování dat, které pocházejí z odvětví informačních technologií. Díky nástrojům, jako jsou Pandas a Apache Parquet, lze kombinovat strukturovaná a nestrukturovaná data, což zjednoduší přístup k informacím a snižuje ztráty při analýze, zatímco velké soubory dat z dokumentů nebo projektů CAD (obr. 9.2-10 - obr. 9.2-12) umožňují shromažďovat, analyzovat a předvídat data ve všech fázích životního cyklu projektu.

Big Data mají na stavebnictví transformační dopad a mohou ho ovlivnit mnoha způsoby. Aplikace technologií Big Data přináší výsledky v řadě klíčových oblastí, mezi něž patří např.:

- **Analýza investičního potenciálu** - předpověď ziskovosti a doby návratnosti projektů na základě údajů z předchozích zařízení.
- **Prediktivní údržba** - identifikace pravděpodobných poruch zařízení dříve, než k nim skutečně dojde, což snižuje prostoje.
- **Optimalizace dodavatelského řetězce** - předvídání narušení a zvyšování efektivity logistiky.
- **Analýzy energetické účinnosti** - pomoc při navrhování nízkoenergetických budov.
- **Bezpečnostní monitoring** - použití senzorů a nositelných zařízení ke sledování podmínek na staveništi.

- **Kontrola kvality** - sledování dodržování procesních norem v reálném čase.
- **Řízení pracovních sil** - analýza výkonnosti a prognóza personálních požadavků.

Ve stavebnictví lze jen těžko najít oblast, kde by nebyla poptávka po datové analytice a predikcích. Hlavní výhodou predikčních algoritmů je jejich schopnost samoučení a neustálého zlepšování s přibývajícími daty.

V blízké budoucnosti bude umělá inteligence nejen pomáhat stavitelům, ale bude činit klíčová rozhodnutí - od projektových procesů až po otázky provozu budov.

Více o tom, jak se generují předpovědi a používají modely učení, se dozvíte v další části knihy "Strojové učení a předpověď".

Přechod k plnohodnotné práci s velkými daty vyžaduje změnu samotného přístupu k analytice. Zatímco klasické systémy, o kterých jsme dosud uvažovali, se zaměřovaly na vztahy příčin a následků, analytika velkých dat přesouvá své zaměření na hledání statistických vzorců a korelací, které umožňují identifikovat skryté vztahy a předpovídat chování objektů i bez úplného pochopení všech faktorů.

Zpochybnění proveditelnosti velkých dat: korelace, statistika a výběr vzorků dat

Tradičně se konstrukce zakládala na subjektivních hypotézách a osobních zkušenostech. Inženýři předpokládali - s určitou mírou pravděpodobnosti - jak se bude materiál chovat, jaká zatížení konstrukce vydrží a jak dlouho projekt vydrží. Tyto předpoklady se ověřovaly v praxi, často za cenu času, zdrojů a budoucích rizik.

S příchodem velkých dat se tento přístup dramaticky mění: rozhodnutí se již nedělají na základě intuitivního tušení, ale jako výsledek analýzy rozsáhlých souborů dat. Konstrukce postupně přestává být uměním intuice a stává se přesnou vědou předvídání.

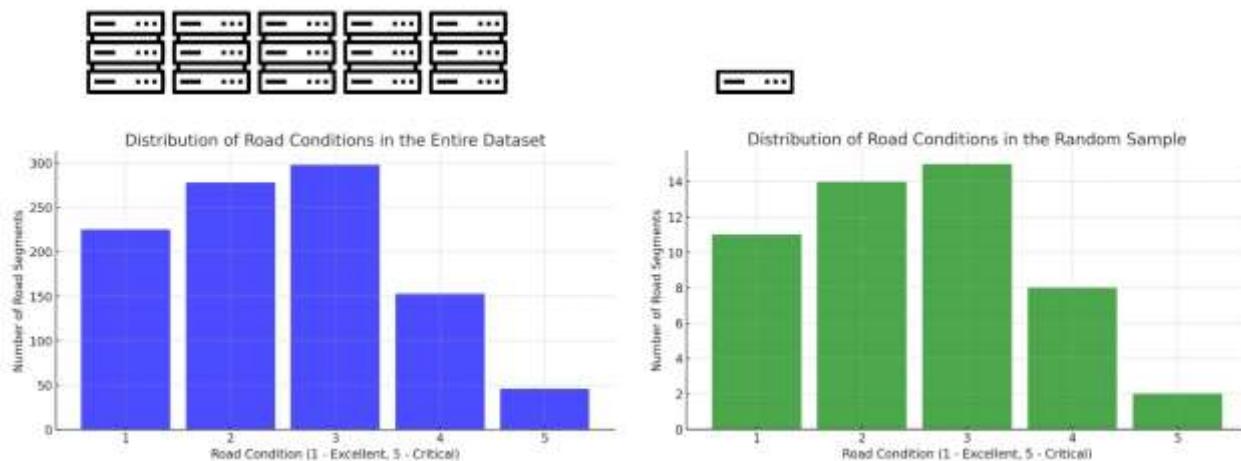
Přechod k myšlence využití velkých dat nevyhnutelně vyvolává důležitou otázku: jak rozhodující je množství dat a kolik informací je skutečně potřeba pro spolehlivou prediktivní analýzu? Rozšířený názor, že "čím více dat, tím vyšší přesnost", se v praxi nevždy ukazuje jako statisticky platný.

Již v roce 1934 dokázal statistik Jerzy Neumann [145], že klíčem k přesnosti statistické inference není ani tak množství dat, jako spíše jejich reprezentativnost a náhodnost výběru.

To platí zejména ve stavebnictví, kde se shromažďuje velké množství dat pomocí senzorů, skenerů, kamer, dronů a dokonce i víceformátových modelů CAD, což zvyšuje riziko vzniku slepých míst, odlehčitých hodnot a zkreslení dat.

Podívejme se na příklad monitorování stavu povrchu vozovky. Kompletní soubor dat všech úseků silnic může mít velikost X GB a jeho zpracování může trvat přibližně jeden den. Současně by náhodný vzorek zahrnující pouze každý 50. úsek silnice zabral pouze X/50 GB a byl by zpracován za půl hodiny, přičemž

by poskytoval podobnou přesnost odhadů pro určité výpočty (obrázek 9.1-1).



Obrázek 9.1-1 Histogramy stavu vozovky: úplný soubor dat a náhodný výběr vykazují shodné výsledky.

Klíčem k úspěšné analýze dat tedy často nemusí být množství dat, ale reprezentativnost vzorku a kvalita použitých metod zpracování. Přechod k náhodnému výběru vzorků a selektivnějšímu přístupu vyžaduje změnu myšlení ve stavebnictví. Historicky se společnosti řídily logikou "čím více údajů, tím lépe" v domnění, že pokrytí všech možných ukazatelů maximalizuje přesnost.

Tento přístup připomíná oblíbený omyl z oblasti projektového řízení: "čím více odborníků přilákám, tím bude práce efektivnější". Stejně jako u lidských zdrojů je však důležitější kvalita a nástroje než kvantita. Bez zohlednění vzájemných vztahů (korelací) mezi daty nebo účastníky projektu může zvyšování objemu vést pouze k šumu, zkreslení, duplicitě a zbytečnému plýtvání.

Nakonec se často ukáže, že je mnohem produktivnější mít k dispozici menší, ale kvalitativně připravený soubor dat, který je schopen vytvářet stabilní a rozumné předpovědi, než se spoléhat na masivní, ale chaotické informace obsahující mnoho protichůdných signálů.

Nadměrný objem dat nejenže nezaručuje větší přesnost, ale může také vést ke zkresleným závěrům v důsledku přítomnosti šumu, nadbytečných rysů, skrytých korelací a irrelevantních informací. Za takových okolností se zvyšuje riziko nadměrného přizpůsobení modelů a snižuje se spolehlivost analytických výsledků.

Ve stavebnictví je hlavní výzvou při práci s velkými daty určení optimálního množství a kvality dat. Například při sledování stavu betonových konstrukcí může použití tisíců senzorů a sběr informací každou minutu zahlit systém ukládání a analýzy. Pokud však provedete korelační analýzu a vyberete 10% nejvíce informativních senzorů, můžete získat téměř identickou přesnost předpovědí, přičemž vynaložíte mnohonásobně, někdy i desítky a stovkykrát, méně prostředků.

Použití menší podmnožiny dat snižuje objem potřebného uložiště i dobu zpracování, což významně snižuje náklady na ukládání a analýzu dat a často činí z náhodného výběru ideální řešení pro prediktivní

analýzu, zejména u velkých infrastrukturních projektů nebo při práci v reálném čase. V konečném důsledku není efektivita stavebních procesů dána množstvím shromážděných dat, ale kvalitou jejich analýzy. Bez kritického přístupu a pečlivé analýzy mohou data vést k nesprávným závěrům.

Po určitém množství dat přináší každá nová jednotka informací stále méně užitečných výsledků. Místo nekonečného shromažďování informací je důležité zaměřit se na jejich reprezentativnost a metody analýzy (obrázek 9.2-2).

Tento jev dobře popsal Allen Wallis [146], který ilustruje použití statistických metod na příkladu testování dvou alternativních konstrukcí střel amerického námořnictva.

Námořnictvo testovalo dvě alternativní konstrukce střel (A a B) sérií párových střel. V každém kole obdrží střela A 1 nebo 0 podle toho, zda je její výkon lepší nebo horší než výkon střely B, a naopak. Standardní statistický přístup zahrnuje provedení pevného počtu pokusů (např. 1000) a určení vítěze na základě procentuálního rozdělení (např. pokud A získá 1 ve více než 53% případů, je považován za nejlepší). Když Allen Wallis diskutoval o takovém problému s kapitánem (námořnictva) Garrettem L. Schuylerem, kapitán namítl, že takový test, abychom citovali Allenovu historku, může být k ničemu. Kdyby byl na místě moudrý a zkušený ordonanční důstojník, jako byl Schuyler, viděl by po prvních několika stech [výstřelů], že experiment není třeba ukončit, ať už proto, že nová metoda je zjevně horší, nebo proto, že je zjevně lepší než to, v co se doufalo [146].

- Vládní statistická výzkumná skupina USA na Kolumbijské univerzitě, období druhé světové války.

Tento princip se hojně využívá v různých průmyslových odvětvích. Například v medicíně se klinické zkoušky nových léků provádějí na náhodných vzorcích pacientů, což umožňuje získat statisticky významné výsledky bez nutnosti testovat lék na celé populaci lidí žijících na této planetě. V ekonomii a sociologii se provádějí reprezentativní průzkumy, které odrážejí názory společnosti, aniž by bylo nutné dotazovat všechny obyvatele země.

Stejně jako vlády a výzkumné organizace provádějí průzkumy v malých populacích, aby pochopily obecné společenské trendy, mohou společnosti ve stavebnictví využívat náhodné vzorky dat k efektivnímu sledování a vytváření prognóz pro řízení projektů (obrázek 9.1-1).

Velká data mohou změnit přístup k sociálním vědám, ale nenahradí statistický zdravý rozum [147].

- Thomas Landsall-Welfair, "Forecasting the nation's current mood", Significance v. 9(4), 2012 .

Z hlediska úspory zdrojů je při sběru dat pro budoucí předpovědi a rozhodování důležité odpovědět na

otázku: Má smysl vynakládat značné prostředky na sběr a zpracování obrovských souborů dat, když lze použít mnohem menší a levnější testovací soubor dat, který lze postupně rozširovat? Efektivita náhodného výběru vzorků ukazuje, že společnosti mohou snížit náklady na sběr a trénování modelů desítky nebo dokonce tisícekrát, pokud zvolí metody sběru dat, které nevyžadují komplexní pokrytí, ale přesto poskytují dostatečnou přesnost a reprezentativnost. Tento přístup umožňuje i malým společnostem dosahovat výsledků srovnatelných s velkými korporacemi za použití výrazně menších zdrojů a objemů dat, což je důležité pro společnosti, které chtějí optimalizovat náklady a urychlit informované rozhodování s využitím malých zdrojů. V následujících kapitolách prozkoumejte příklady analytiky a prediktivní analytiky založené na veřejných datových souborech s využitím nástrojů big data.

Velká data: analýza dat ze souboru milionů stavebních povolení v San Francisku

Práce s otevřenými soubory dat poskytuje jedinečnou příležitost uplatnit v praxi zásady, o nichž jsme hovořili v předchozích kapitolách: uvážlivý výběr prvků, reprezentativní výběr vzorků, vizualizace a kritická analýza. V této kapitole se budeme zabývat tím, jak lze pomocí otevřených dat zkoumat složité jevy, jako je stavební činnost ve velkém městě - konkrétně více než milion záznamů o stavebních povoleních v San Franciscu.

Veřejně dostupné údaje o více než milionu stavebních povolení (obrázek 9.1-2) (záznamy ve dvou datových souborech ve formátu CSV) z "San Francisco Department of Buildings". [148] nám umožňují použít surovou tabulku CSV -tabulku nejen k analýze stavební činnosti ve městě, ale také ke kritické analýze nedávných trendů a historie sanfranciského stavebnictví za posledních 40 let, od roku 1980 do roku 2019.

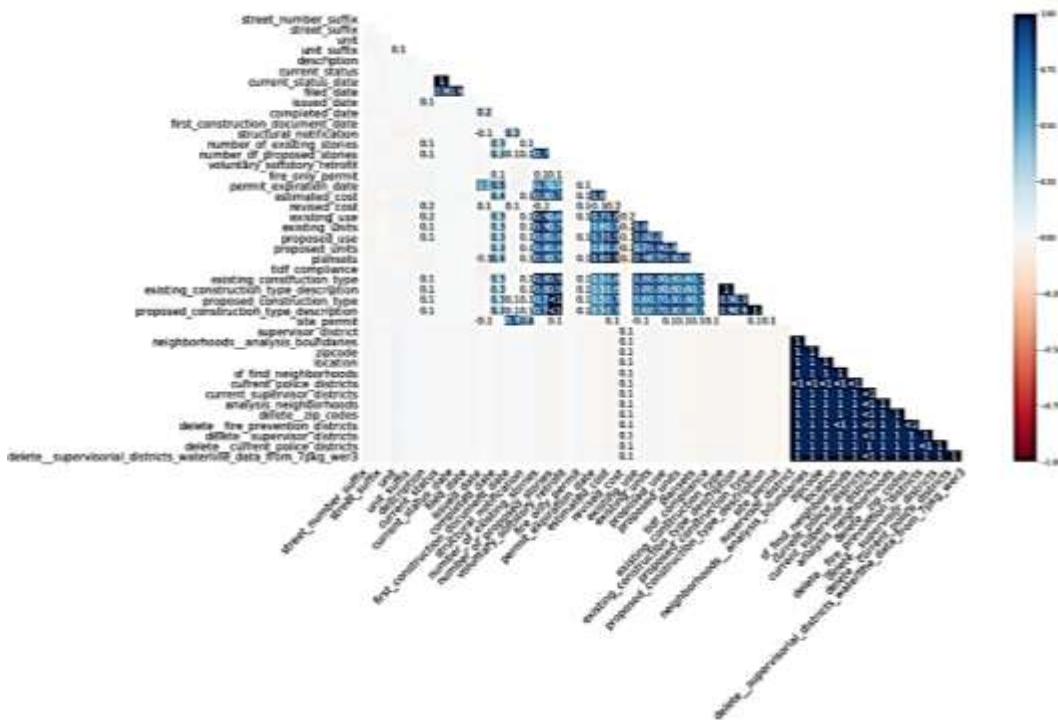
Příklady kódu použité k vytvoření vizualizace datové sady (obrázky 9.1-3- Obrázky 9.1-8), stejně jako vizuální grafy s kódem, vysvětlivkami a komentáři, lze nalézt na platformě Kaggle po vyhledání "San Francisco. Sektor stavebnictví 1980-2019". [149].

The screenshot displays two tables of data from the San Francisco Department of Buildings. The top table is titled 'Building Permits on or after January 1, 2013' and shows a count of 1.137695e+06 permits. The bottom table is titled 'Building Permits before January 1, 2013' and shows a subset of permits with various attributes like permit creation date, description, current status, and location.

permit_creation_date	description	current_status	current_status_date	filed_date	issued_date	completed_date
07/01/1998	repair stucco	complete	07/07/1998	07/01/1998	07/01/1998	07/07/1998
12/13/2004	reroofing	expired	01/24/2006	12/13/2004	12/13/2004	NaN
02/18/1992	install auto fire spks.	complete	06/29/1992	02/18/1992	03/18/1992	06/29/1992

permit_number	permit_expiration_date	estimated_cost	revised_cost	existing_use	zipcode	location
362780	9812394	11/01/1998	780.0	NaN	1 family dwelling	94129.0 (-127.7936468760498, -122.4322641443574)
570817	200412131233	06/13/2005	9000.0	9000.0	apartments	94127.0 (-127.729258510008388, -122.4644245687462)
198411	9202396	09/18/1992	9000.0	NaN	apartments	94111.0 (-127.79806002552974, -122.39593224461805)

Obr. 9.1-2 Datové sady obsahují informace o vydaných stavebních povoleních s různými atributy objektu.

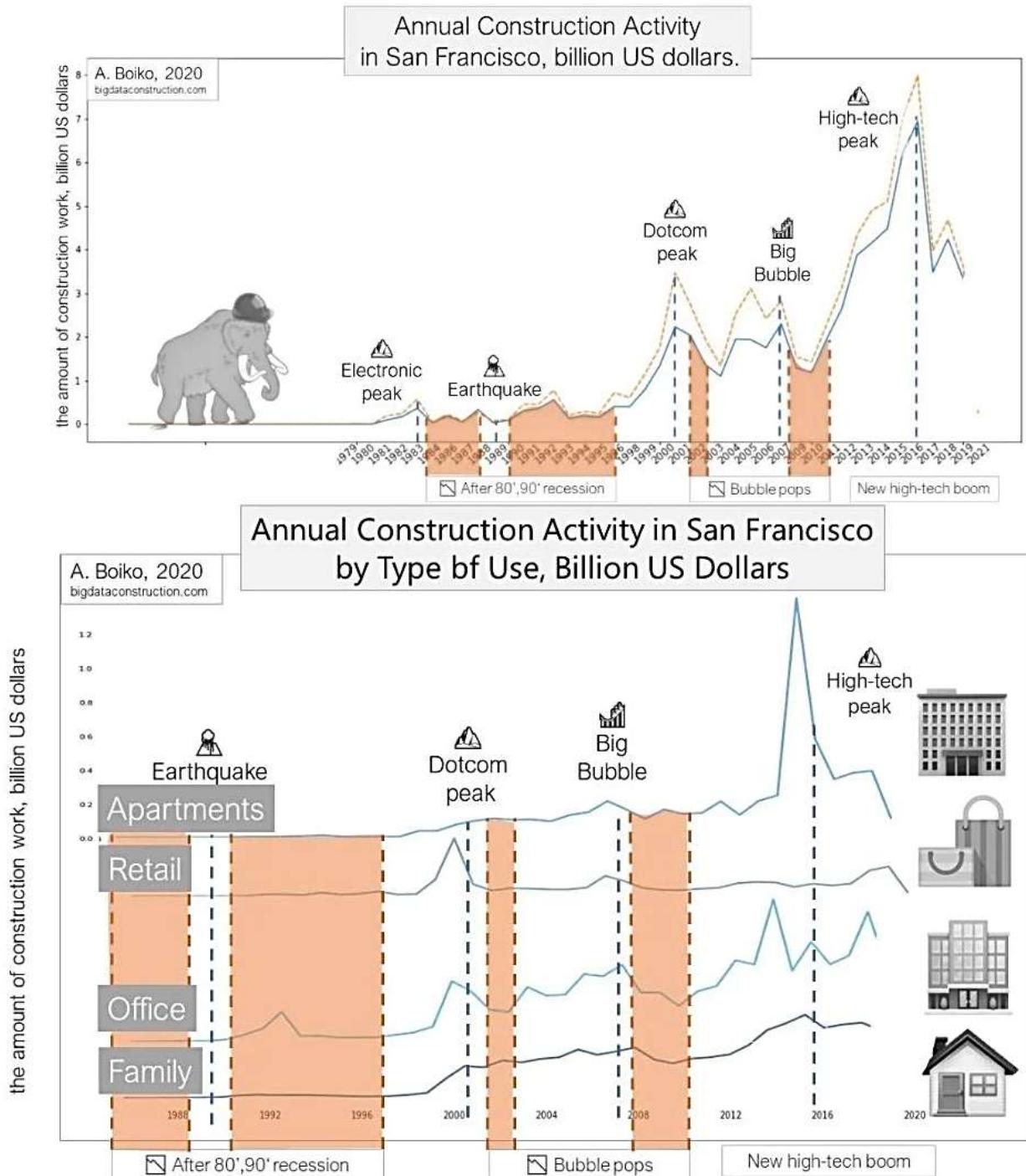


Obr. 9.1-3 Tepelná mapa (Pandas a Seaborn), která vizualizuje všechny atributy datového souboru a pomáhá identifikovat vztahy mezi dvojicemi atributů.

Z tabulky, kterou poskytl sanfranciský odbor stavebnictví (obrázek 9.1-2), nejsou patrné žádné trendy ani závěry. Suchá čísla v tabulkové podobě nejsou podkladem pro rozhodování. Aby byla data vizuálně srozumitelná, jak je podrobně popsáno v kapitolách o vizualizaci dat, měla by být vizualizována pomocí různých knihoven, o nichž se pojednává v sedmé části knihy na téma "ETL a vizualizace výsledků v podobě grafů".

Na základě analýzy dat, která byla provedena pomocí knihoven Pandas DataFrame a Python visualisation hodnotě 1 137 695 povolení [148], můžeme dojít k závěru, že stavební činnost v San Franciscu je úzce spjata s ekonomickými cykly, zejména v rychle se rozvíjejícím technologickém průmyslu Silicon Valley (obrázek 9.1-4).

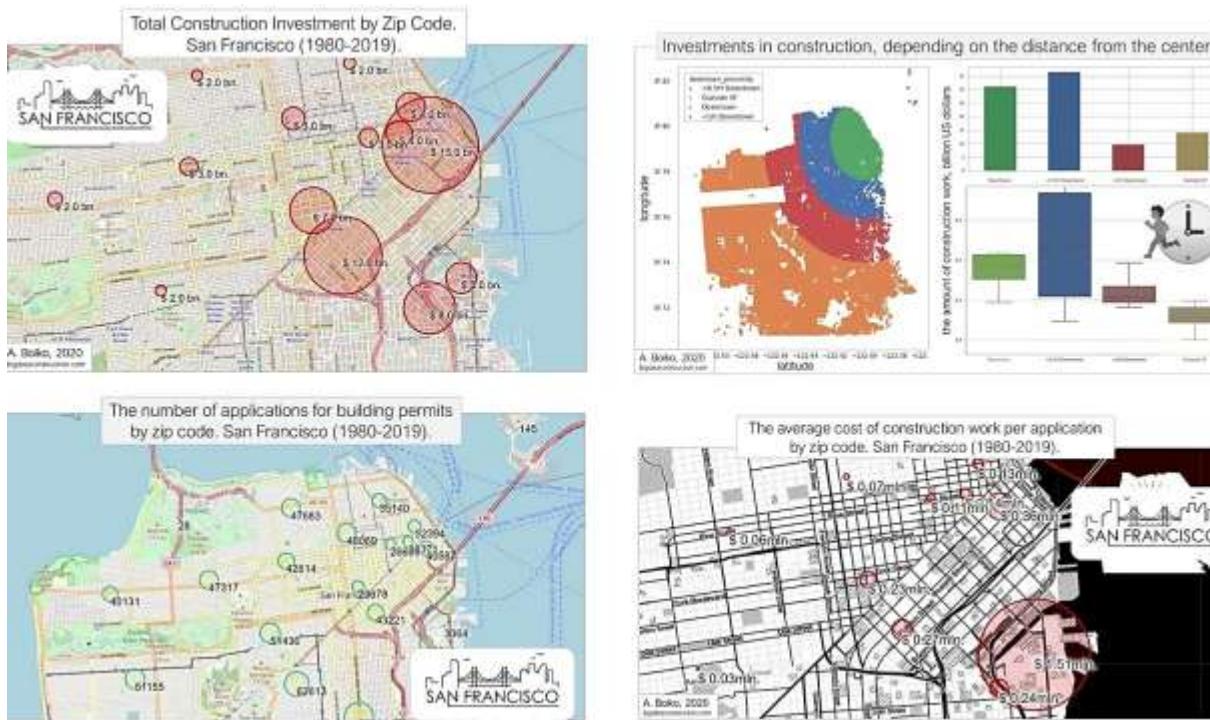
Hospodářské konjunktury a propady mají významný vliv na počet a hodnotu stavebních projektů. Například první vrchol stavební činnosti se shodoval s boomem elektroniky v polovině 80. let (použito Pandas a Matplotlib) a následné vrcholy a poklesy souvisely s bublinou dot-com a technologickým boomem posledních let.



Obr. 9.1-4 V sektoru nemovitostí v San Francisku korelují investice s technologickým rozvojem Silicon Valley.

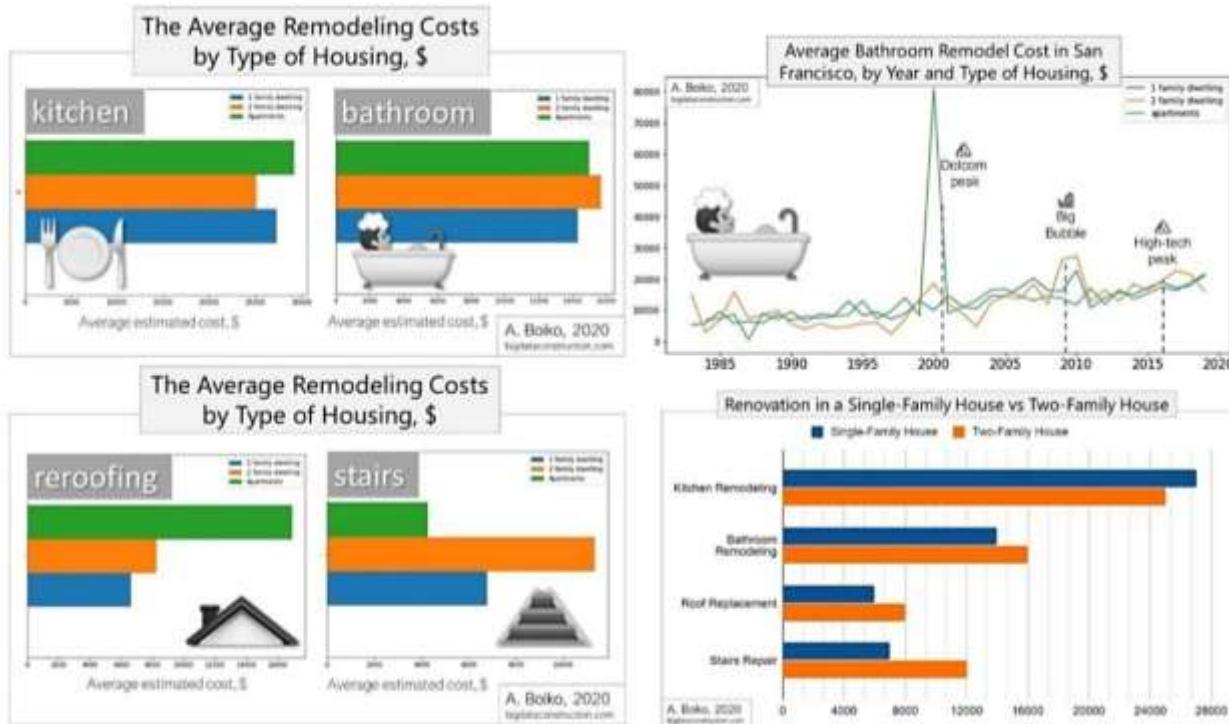
Analýza dat naznačuje, že v San Franciscu se většina 91,5 miliardy dolarů investovaných do výstavby a přestavby za posledních deset let - téměř 75% - soustředí v centru města (obrázek 9.1-5 - použito Pandas a vizualizační knihovna Folium) a v okruhu 2 km od centra, což odráží vyšší hustotu investic v těchto centrálních zónách.

Průměrné náklady na stavební povolení se v jednotlivých čtvrtích značně liší, přičemž žádosti v centru města jsou třikrát dražší než žádosti mimo centrum města, a to kvůli vyšším nákladům na pozemky, pracovní sílu, materiály a přísným stavebním předpisům, které vyžadují dražší materiály pro zlepšení energetické účinnosti.



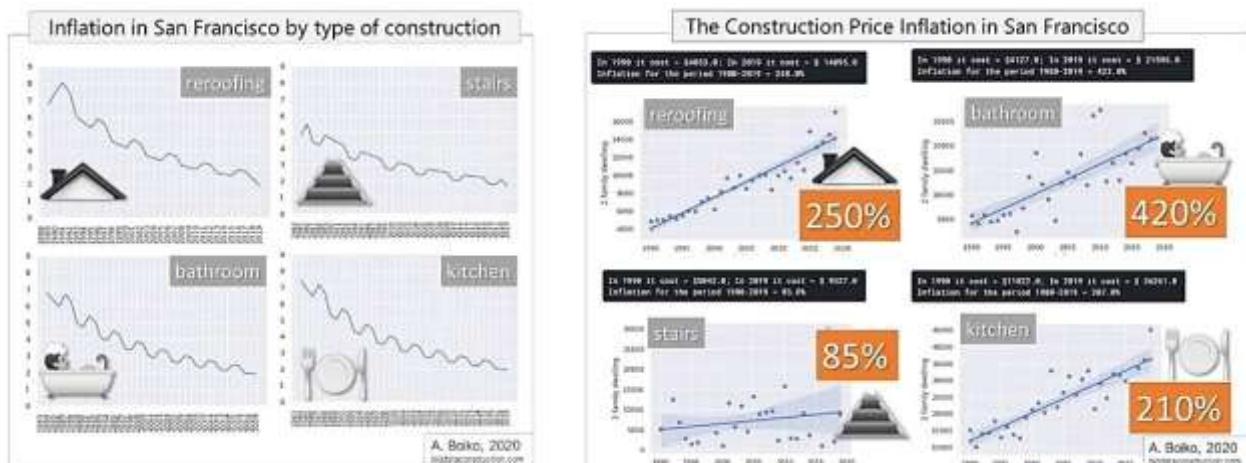
Obrázek 9.1-5 V San Franciscu je 75% stavebních investic (91,5 miliardy USD) soustředěno v centru města.

Soubor dat také umožnuje výpočet průměrných cen oprav nejen podle typu domu, ale také podle městských čtvrtí a jednotlivých adres (poštovních směrovacích čísel). Dynamika nákladů na opravy domů v San Francisku vykazuje odlišné trendy pro různé typy oprav a bydlení (obrázek 9.1-6 - použity Pandas a Matplotlib). Rekonstrukce kuchyně jsou znatelně dražší než rekonstrukce koupelny: průměrná rekonstrukce kuchyně v rodinném domě stojí přibližně 28 000 USD, zatímco v případě dvougeneračního domu 25 000 USD.



Obrázek 9.1-6 Ve SF stojí renovace kuchyně téměř dvakrát také co renovace koupelny a majitelé domů si musí každý měsíc po dobu 15 let odkládat \$ 350, aby pokryli náklady na větší opravy domu.

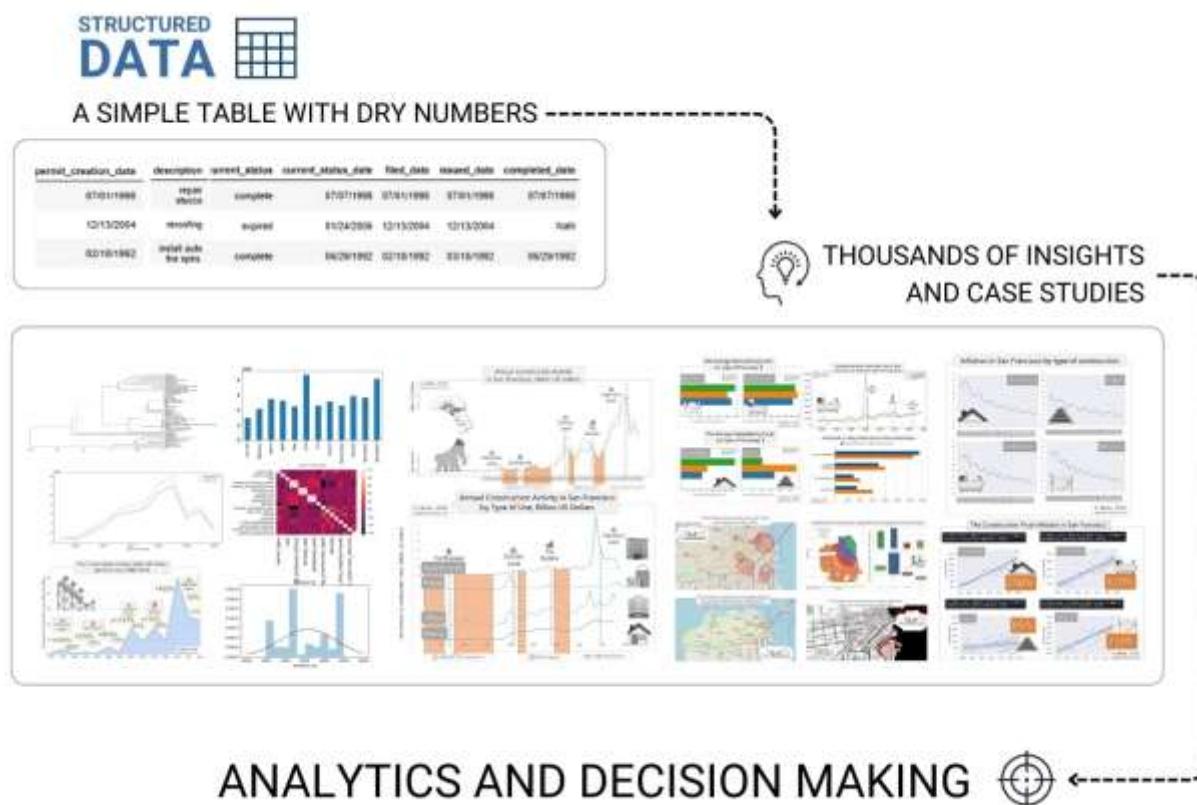
Inflaci stavebních nákladů v San Franciscu v průběhu let lze vysledovat na základě analýzy údajů seskupených podle typu bydlení a roku (obrázek 9.1-7 - použity Pandas a Seaborn), který ukazuje trvalý nárůst průměrných nákladů na opravy od roku 1990 a odhaluje krátkodobé tříleté cykly v nákladech na opravy vícebytových budov.



Obrázek 9.1-7 Od roku 1980 do roku 2019 vzrostly náklady na renovaci koupelny ve SF pětinásobně, zatímco renovace střechy a kuchyně se ztrojnásobily a renovace schodišť se zvýšily pouze o 85%.

Studie veřejných údajů Stavebního odboru San Francisca (obrázek 9.1-3) ukazuje, že stavební náklady ve městě jsou velmi proměnlivé a často nepředvídatelné, ovlivněné řadou faktorů. Mezi tyto faktory patří hospodářský růst, technologické inovace a jedinečné požadavky různých typů bydlení.

V minulosti takové analýzy vyžadovaly hluboké znalosti programování a analytiky. S příchodem nástrojů LLM- se však tento proces stal přístupným a srozumitelným širokému okruhu odborníků ve stavebnictví, od inženýrů v projekčních odděleních až po vrcholový management.



Obr. 9.1-8 Přechod na vizuálně srozumitelná data umožňuje automatizované rozhodování díky rozpoznávání skrytých vzorců.

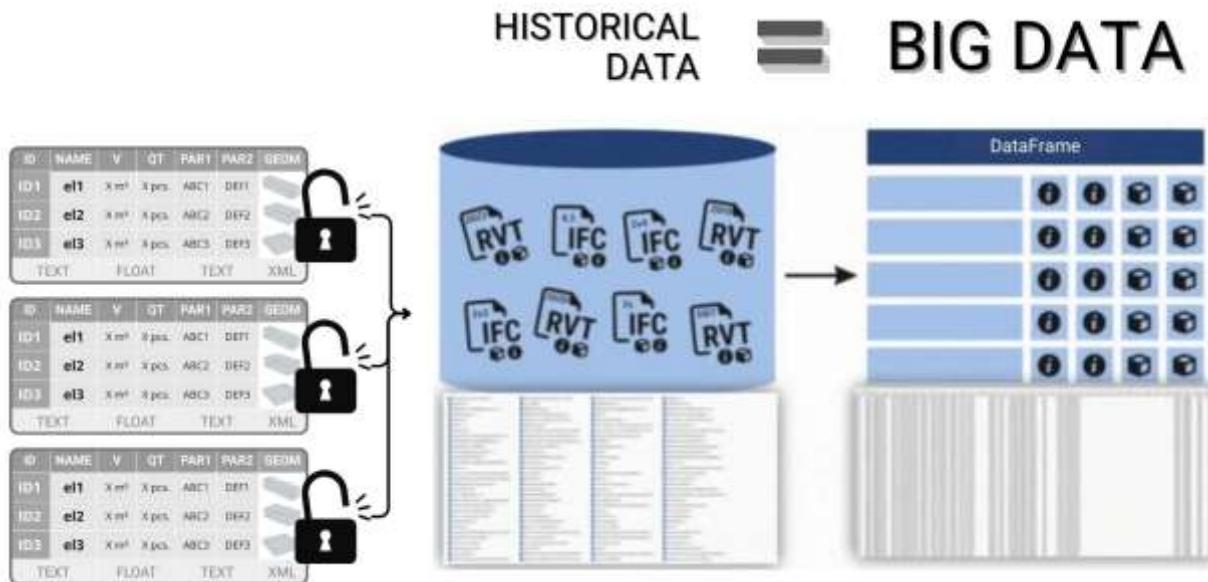
Stejně jako jsme analyzovali data z tabulkového souboru dat "San Francisco Building Authority", můžeme vizualizovat a analyzovat jakýkoli soubor dat - od obrázků a dokumentů až po data internetu věcí, nebo data z odvozených databází CAD.

Příklad velkých dat na základě dat CAD (BIM).

V následujícím příkladu budeme analyzovat rozsáhlý soubor dat s využitím dat z různých nástrojů CAD (BIM). Ke shromáždění a vytvoření rozsáhlého souboru dat byl použit specializovaný automatizovaný webový crawler (skript), který byl nakonfigurován tak, aby automaticky vyhledával a shromažďoval soubory návrhů z webových stránek nabízejících bezplatné architektonické modely ve formátech RVT a IFC. Během několika dní crawler úspěšně našel a stáhl 4 596 souborů IFC a 6 471 souborů RVT a

156 024 souborů DWG[149].

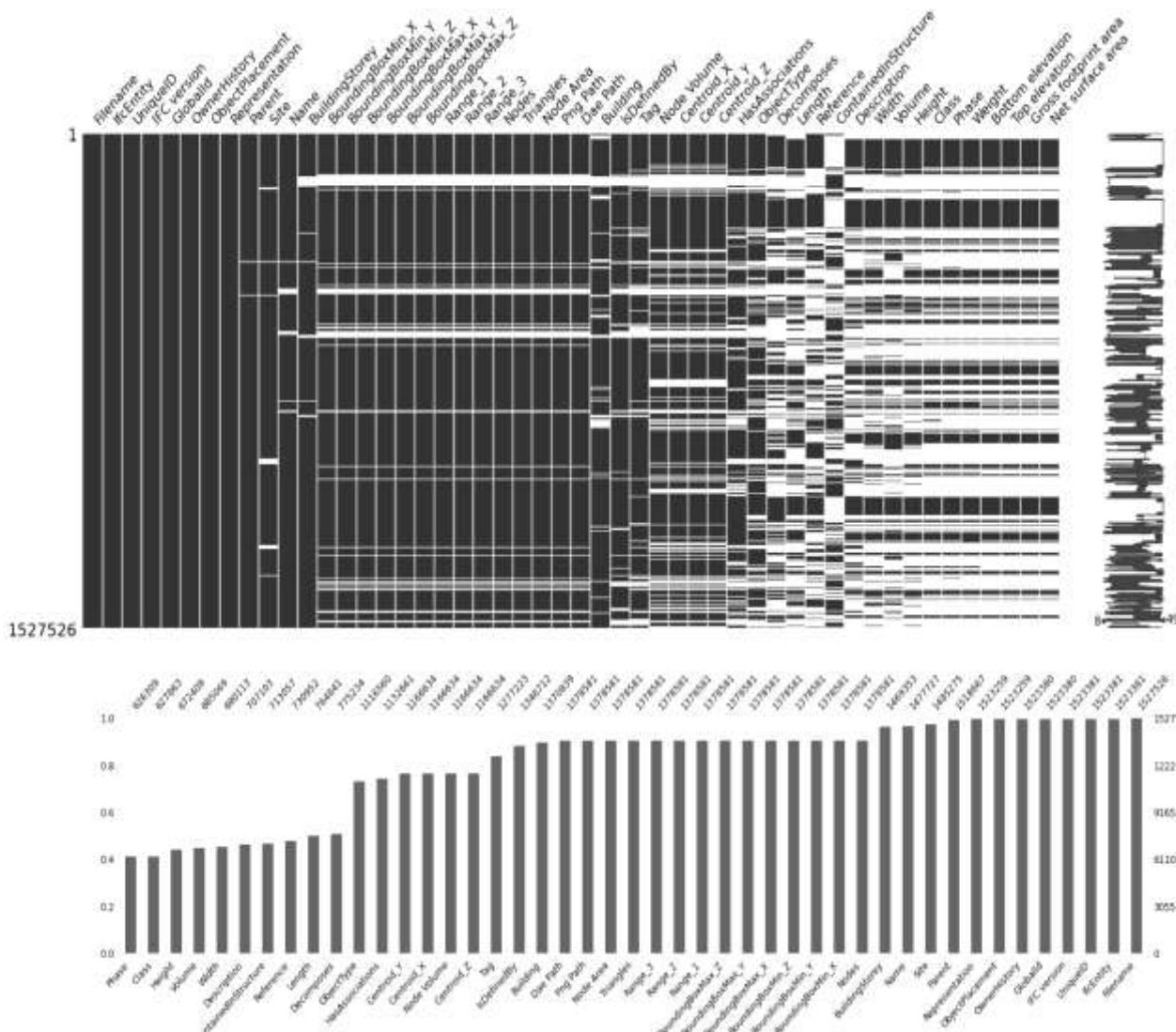
Po shromáždění projektů ve formátech RVT a IFC různých verzí a jejich převedení do strukturovaného formátu CSV pomocí bezplatného reverzního inženýrství SDK bylo téměř 10 tisíc projektů RVT a IFC shromážděno do jednoho velkého souboru tabulky Apache Parquet a nahráno do Pandas DataFrame pro analýzu (obrázek 9.1-9).



Obrázek 9.1-9 Strukturovaná data data projektů umožňuje spojit libovolný počet projektů do jedné dvouozměrné tabulky.

Údaje z této rozsáhlé sbírky obsahují následující informace: soubor souborů IFC obsahuje přibližně 4 miliony entit (řádků) a 24 962 atributů (sloupců) a soubor souborů RVT, který se skládá z přibližně 6 milionů entit (řádků), obsahuje 27 025 různých atributů (sloupců).

Tyto soubory informací (obrázek 9.1-10) zahrnují miliony prvků, pro každý z nich byly dodatečně získány souřadnice geometrie Bounding Box (obdélník vymezující hranice objektu v projektu) a přidány do společné tabulky - byly vytvořeny souřadnice geometrie Bounding Box (obdélník vymezující hranice objektu v projektu) a obrázky každého prvku ve formátu PNG a geometrie v otevřeném formátu XML - DAE (Collada).



Obr. 9.1-10 Sabset 1,5 milionu prvků a vizualizace (knihovna missingno) obsazení prvních 100 atributů ve formě histogramu.

Získali jsme tak veškeré informace o desítkách milionů prvků ze 4 596 projektů IFC a 6 471 projektů RVT, kde byly všechny atributy-vlastnosti všech entitních prvků a jejich geometrie (Bounding Box) převedeny do strukturované podoby jediné tabulky (DataFrame) (obr. 9.1-10 - údaje o populacích datových rámců se zobrazují jako histogramy).

Histogramy (obr. 9.1-10, obr. 9.2-6, obr. 9.2-7) vykreslené během procesu analýzy umožňují rychlé posouzení hustoty dat a četnosti výskytu hodnot ve sloupcích. To umožňuje první náhled na rozložení prvků, přítomnost odlehčitelných hodnot a potenciální užitečnost jednotlivých atributů při analýze a sestavování modelů strojového učení.

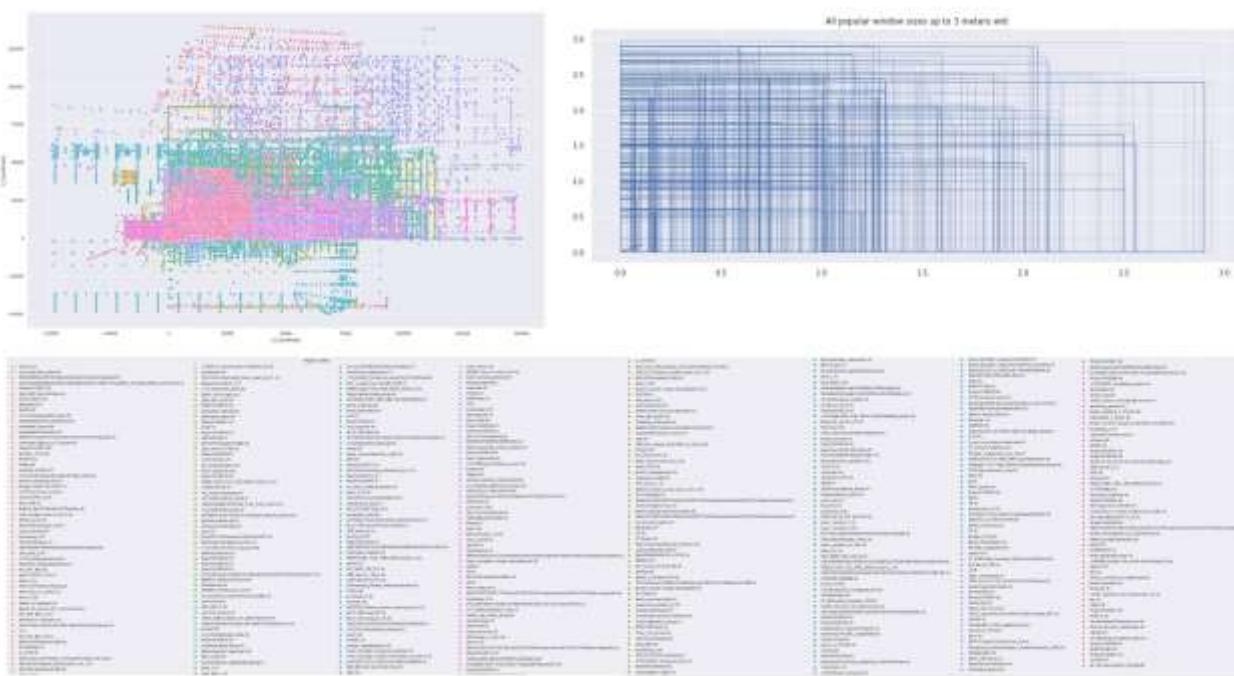
Jedním z příkladů praktického využití této sady dat (obr. 9.1-10) je projekt "5000 IFC a RVT ". [149], který je k dispozici na platformě Kaggle. Představuje Jupyter Notebook s kompletním řešením Pipeline:

od předzpracování a analýzy dat až po vizualizaci výsledků pomocí knihoven Pythonu - pandas, matplotlib, seaborn, folium a dalších (obr. 9.1-11).



Obrázek 9.1-11 Příklady analýzy dat z formátů CAD (BIM) pomocí vizualizačních knihoven Pythonu a knihovny pandas.

Na základě metainformací je možné určit, ve kterých městech byly určité projekty realizovány, a zobrazit je na mapě (např. pomocí knihovny folium). Časová razítka v datech navíc umožňují zkoumat vzorce, kdy byly soubory uloženy nebo upraveny: podle dne v týdnu, denní doby a měsíce.



Obr. 9.1-12 Vizualizace geometrické polohy všech sloupů a rozměry všech oken do 3 metrů v projektech ze seznamu v dolní části grafu.

Geometrické parametry v podobě Bounding Boxu získané z modelů se rovněž hodí k souhrnným

analýzám. Například na obrázku 9.1-12 jsou zobrazeny dva grafy: levý ukazuje rozložení vzdáleností mezi sloupci pro všechny projekty vzhledem k nulovému bodu a pravý zobrazuje rozměry všech oken do výšky 3 m ve vzorku desítek tisíc okenních prvků (po seskupení celého souboru dat podle parametru "Category" s hodnotou "OST_Windows", "IfcWindows").

Kód analýzy Pipeline pro tento příklad a samotný soubor dat jsou k dispozici na webových stránkách Kaggle pod názvem "5000 IFC and RVT | DataDrivenConstruction.io projects" [149]. Tuto hotovou Pipeline spolu s datovou sadou lze zkopirovat a spustit zdarma online na Kaggle nebo offline v některém z populárních IDE: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse s pluginem PyDev, Thonny, Wing IDE, IntelliJ IDEA s pluginem Python, JupyterLab nebo populární online nástroje Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Analytické poznatky získané zpracováním a studiem obrovského množství strukturovaných dat budou hrát klíčovou roli v rozhodovacích procesech ve stavebnictví.

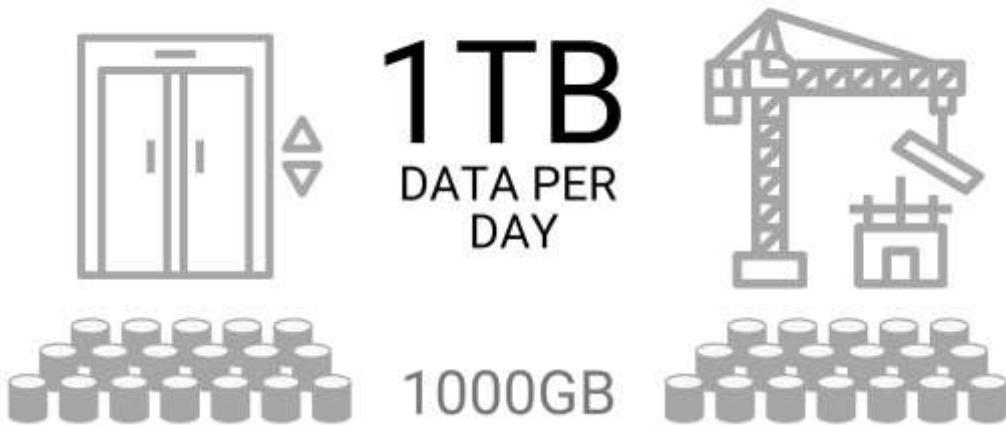
Díky takové analýze informací na základě minulých projektů mohou odborníci efektivně předvídat například požadavky na materiál a pracovní sílu a optimalizovat konstrukční řešení ještě před zahájením stavby

Zatímco však projektová data nebo stavební povolení jsou relativně statické informace, které se mění poměrně pomalu, samotný proces výstavby se rychle nasycuje nejrůznějšími senzory a zařízeními internetu věcí: kamerami, automatizovanými monitorovacími systémy, které přenášejí data v reálném čase - to vše mění staveniště v dynamické digitální prostředí, kde je třeba data analyzovat v reálném čase.

Internet věcí a chytré smlouvy

Internet věcí Internet věcí představuje novou vlnu digitální transformace, v níž každé zařízení získává vlastní IP adresu a stává se součástí globální sítě. Internet věcí je koncept, který zahrnuje připojení fyzických objektů k internetu za účelem sběru, zpracování a přenosu dat. Ve stavebnictví to znamená možnost řídit stavební procesy v reálném čase, minimalizovat plýtvání materiálem, předvídat opotřebení zařízení a automatizovat rozhodování.

Podle článku CFMA "Příprava na budoucnost s propojenou výstavbou" [150] projde stavebnictví v příštím desetiletí zásadní digitální transformací, která vyvrcholí konceptem Connected Construction - plně integrovaného a automatizovaného staveniště.



Obr. 9.1-13 Zařízení IoT nebo zařízení pro přenos dat ze staveniště mohou denně produkovat a přenášet terabajty dat.

Digitální staveniště znamená, že všechny prvky výstavby - od plánování a logistiky až po provádění prací a kontrolu kvality na staveništi pomocí pevných kamer a kvadrokoptér - budou integrovány do jediného dynamického digitálního ekosystému. Již dříve, v 7. části této knihy, jsme se zabývali možnostmi Apache NiFi (obrázek 7.4-5), bezplatného nástroje s otevřeným zdrojovým kódem, který umožňuje streamování dat v reálném čase - od sběru z různých zdrojů až po přenos do úložišť nebo analytických platform.

Údaje o postupu výstavby, spotřebě materiálu, stavu zařízení a bezpečnosti budou v reálném čase přenášeny do analytických systémů (obr. 9.1-13). To umožňuje předvídat potenciální rizika, okamžitě reagovat na odchylky a optimalizovat procesy na stavbě. Mezi klíčové součásti digitálního staveniště patří např:

- IoT -senzory - sledování parametrů prostředí, monitorování stavebních zařízení a kontrola pracovních podmínek.
- Digitální dvojčata - virtuální modely budov a infrastruktury, které umožňují předvídat možné odchylky a předcházet chybám.
- Automatizované logistické systémy - řízení dodavatelského řetězce v reálném čase pro snížení prostojů a nákladů.
- Robotické stavební systémy - využití autonomních strojů k provádění rutinních i nebezpečných úkolů.

Robotizace využívání internetu věcí a koncepce digitálního staveniště (Connected Site) nejen zvýší efektivitu a sníží náklady, ale také zahájí novou éru bezpečnosti, udržitelné výstavby a prediktivního řízení projektů.

Tagy RFID (Radio Frequency Identification) jsou také jednou z nejdůležitějších součástí internetu věcí. Používají se k identifikaci a sledování materiálů, strojů a dokonce i pracovníků na staveništi, čímž zvyšují transparentnost a kontrolu nad zdroji projektu.

RFID -technologie se používá k automatickému rozpoznávání objektů pomocí rádiových signálů.

Skládá se ze tří klíčových prvků:

- RFID - štítky (pasivní nebo aktivní) - obsahují jedinečný identifikátor a jsou připevněny na materiály, nástroje nebo stroje.
- Skenery jsou zařízení, která čtou informace ze štítků a přenášejí je do systému.
- Centralizovaná databáze - uchovává informace o poloze, stavu a pohybu objektů.

Použití RFID ve stavebnictví:

- Automatická evidence materiálu - štítky na hotových betonových směsích, výztuži nebo obalech sendvičových panelů umožňují kontrolu zásob a zabraňují krádežím.
- Kontrola práce personálu - RFID - zaměstnanecké odznaky zaznamenávají časy začátku a konce směny, čímž zajišťují evidenci pracovní doby.
- Monitorování zařízení - RFID - systém sleduje pohyb zařízení, zabraňuje prostojům a zvyšuje efektivitu logistiky.

Tuto sadu technologií doplňují inteligentní smlouvy založené na blockchainu, které automatizují platby, kontrolu dodávek a dodržování smluv bez nutnosti zprostředkovatelů, čímž snižují riziko podvodů a zpoždění.

Vzhledem k absenci společného datového modelu jsou dnes chytré smlouvy pouhým kódem, na kterém se účastníci dohodnou. Díky datově orientovanému přístupu je však možné vytvořit společný model parametrů smlouvy, zakódovat jej do blockchainu a automatizovat plnění podmínek.

Například v systému řízení dodavatelského řetězce může inteligentní smlouva sledovat doručení zásilky pomocí senzorů IoT a RFID a automaticky převést platbu, jakmile dorazí. Podobně na staveništi může chytrá smlouva zaznamenat dokončení pracovní fáze - například instalaci výztuže nebo vylití základů - na základě údajů z dronů nebo stavebních senzorů a automaticky iniciovat další platbu dodavateli bez nutnosti ruční kontroly a papírových potvrzení.

Navzdory novým technologiím a úsilí mezinárodních standardizačních organizací však komplikuje prostředí internetu věcí množství konkurenčních norem.

Podle studie společnosti Cisco zveřejněné v roce 2017 [151] se téměř 60% iniciativ v oblasti internetu věcí (IoT) zastaví ve fázi ověření konceptu a pouze 26% společností považuje své projekty v oblasti internetu věcí za plně úspěšné. Třetina dokončených projektů navíc nedosahuje stanovených cílů a ani po realizaci není uznána za úspěšnou.

Jedním z klíčových důvodů je nedostatečná interoperabilita mezi platformami, které zpracovávají data z různých senzorů. V důsledku toho zůstávají data izolovaná v rámci jednotlivých řešení. Alternativou k tomuto přístupu, stejně jako v jiných podobných případech (kterým jsme se věnovali v této knize), je architektura postavená na samotných datech jako primárním aktivu.

Senzory internetu věcí hrají klíčovou roli nejen při sledování technického stavu zařízení, ale také při prediktivní analýze, která snižuje rizika na staveništi a zlepšuje celkovou výkonnost procesů předvídaním poruch a odchylek.

Data shromážděná senzory IoT a tagy RFID mohou být v reálném čase zpracovávána algoritmy strojového učení, které dokáží odhalit anomálie a předem upozornit inženýry na možné poruchy. Může se jednat o různé situace, od mikrotrhlin v betonových konstrukcích až po nezvyklé pauzy v provozu věžových jeřábů, které indikují technické poruchy nebo porušení předpisů. Kromě toho mohou pokročilé algoritmy analýzy chování zachytit vzorce chování, které mohou indikovat například fyzickou únavu personálu, a zlepšit tak proaktivní řízení bezpečnosti a pohody zaměstnanců na staveništi.

Ve stavebnictví dochází k nehodám a selháním - ať už strojů nebo lidí - jen zřídka náhle. Obvykle jim předcházejí drobné odchylky, které zůstanou nepovšimnutý. Prediktivní analýza a strojové učení umožňují tyto signály odhalit v rané fázi, ještě předtím, než nastanou kritické následky.

Digitální stopu stavebních projektů tvoří dokumenty, projektové soubory a data ze zařízení IoT a štítků RFID, strojové učení z nich však může pomoci získat užitečné informace. S nárůstem objemu dat a demokratizací přístupu k nim získává stavebnictví nové možnosti v oblasti analytiky, prediktivní analýzy a aplikací umělé inteligence.

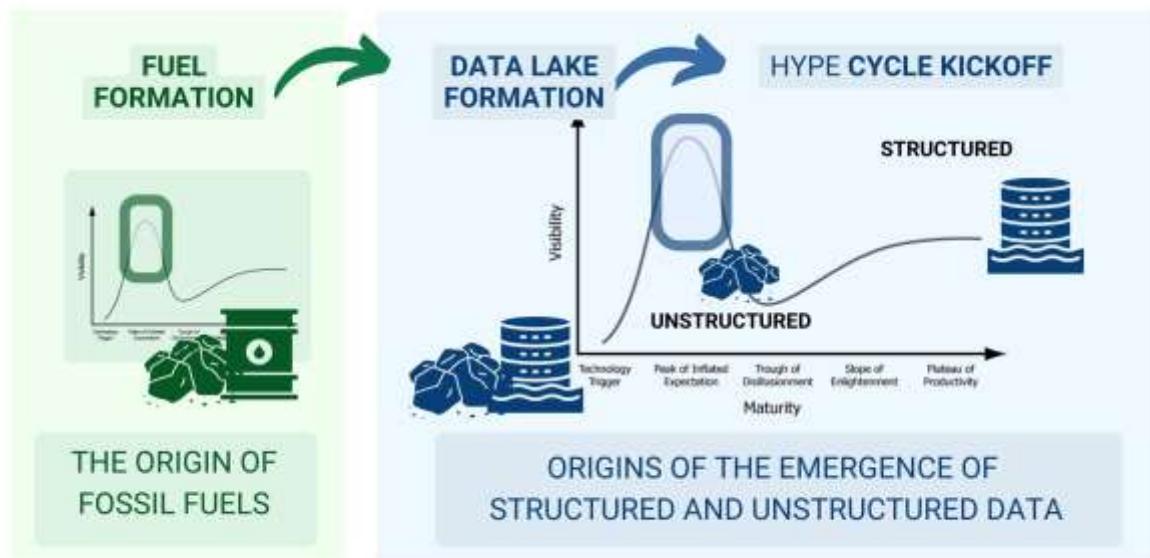


KAPITOLA 9.2.

STROJOVÉ UČENÍ A PŘEDPOVĚDI

Strojové učení a umělá inteligence změní způsob, jakým stavíme.

Databáze různých systémů ve stavebnictví - s jejich nevyhnutelně se rozpadající a stále složitější infrastrukturou - se stávají živou půdou pro budoucí řešení. Firemní servery jsou podobně jako les bohaté na biomasu důležitých informací, často ukrytých pod zemí, v útrobach složek a serverů. Masa dat z různých dnes vytvářených systémů po použití, po pádu na dno serveru a po letech fosilizace - bude v budoucnu pohánět strojové učení a jazykové modely. Interní firemní chaty (např. samostatná instance lokálně nakonfigurovaného ChatGPT, LLaMa, Mistral, DeepSeek) budou postaveny na těchto interních modelech využívajících centralizované úložiště pro rychlé a pohodlné získávání informací a generování potřebných grafů, přehledů a dokumentů.



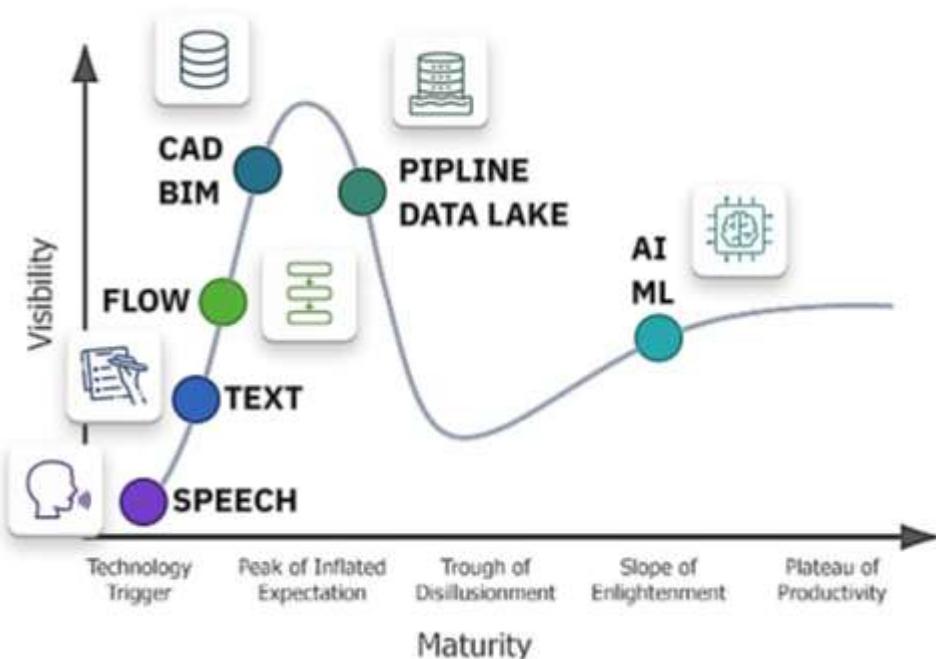
Obr. 9.2-1 Stejně jako se stromy mění v uhlí, tak se i informace pod tlakem času a analytiky časem mění v cennou obchodní energii.

Fosilizace rostlinné hmoty v kombinaci s tlakem a teplotou vytváří homogenní a jedinečně strukturovanou stejnorodou hmotu stromů různých druhů, které žily v různých dobách - dřevěné uhlí [152]. Stejně tak informace zaznamenané na pevných discích v různých formátech a v různých dobách pod tlakem analytických oddělení a teplotou řízení kvality nakonec vytvoří homogenní strukturovanou masu cenných informací (obr. 9.2-1).

Tyto vrstvy (nebo častěji izolované nugety) informací vznikají pečlivou organizací dat zkušenými analytiky, kteří ze zdánlivě dlouho irelevantních dat začnou postupně získávat cenné informace.

V okamžiku, kdy tyto vyspělé datové vrstvy již nejsou jen "vypáleny" v reportech, ale začnou obíhat v podnikových procesech, obohatovat rozhodování a zlepšovat procesy, stává se podnik připraven na další krok - přechod na strojové učení a umělou inteligenci (obr. 9.2-2).

Strojové učení (ML - Machine learning) je třída metod pro řešení problémů umělé inteligence. Algoritmy strojového učení rozpoznávají vzory ve velkých souborech dat a používají je k vlastnímu učení. Každý nový soubor dat umožňuje matematickým algoritmům zlepšovat se a přizpůsobovat se podle získaných informací, což umožňuje neustále zlepšovat přesnost doporučení a předpovědí.



Obr. 9.2-2 Zanikající technologie tvorby dat a aplikace analytických nástrojů otevírá dveře tématu strojového učení.

Jak řekl v roce 2023 v rozhovoru vlivný generální ředitel největšího investičního fondu na světě (který vlastní klíčové podíly v téměř všech největších společnostech vyrábějících stavební software a také ve společnostech, které vlastní největší množství nemovitostí na světě [55]). - strojové učení změní svět stavebnictví.

Umělá inteligence má obrovský potenciál. Změní způsob, jakým pracujeme a žijeme. Umělá inteligence a robotika změní způsob, jakým pracujeme a jakým stavíme, a budeme moci využít umělou inteligenci a robotiku jako prostředek k vytvoření mnohem vyšší produktivity [153].

- Generální ředitel největšího investičního fondu na světě, rozhovor, září 2023.

Strojové učení (ML) funguje na základě zpracování velkého množství dat pomocí statistických technik, které napodobují aspekty lidského myšlení. Většina společností však takové soubory dat nemá, a pokud ano, často nejsou dostatečně označeny. Zde mohou pomoci sémantické technologie a

transferové učení, technika, která umožňuje ML efektivněji pracovat s malým množstvím dat a jejíž proveditelnost byla diskutována v předchozích kapitolách této části.

Podstatou transferového učení je, že místo toho, abyste se každý úkol učili od začátku, můžete využít znalosti získané v příbuzných oborech. Je třeba si uvědomit, že vzory a objevy z jiných odvětví lze přizpůsobit a aplikovat ve stavebnictví. Například metody optimalizace logistických procesů vyvinuté v maloobchodě pomáhají zvýšit efektivitu řízení dodavatelského řetězce ve stavebnictví. Analýzu velkých dat, která se aktivně využívá ve finančníctví, lze aplikovat na prognózování nákladů a řízení rizik ve stavebních projektech. A technologie počítačového vidění a robotiky, které se vyvíjejí v průmyslu, již nacházejí uplatnění v automatické kontrole kvality, monitorování bezpečnosti a řízení zařízení staveniště.

Transferové učení umožňuje nejen urychlit zavádění inovací, ale také snížit náklady na jejich vývoj s využitím již nashromážděných zkušeností z jiných odvětví.

$$\text{labor productivity in construction} = f(\text{AI})$$

Obr. 9.2-3 Umělá inteligence technologie a robotika budou v budoucnu hlavní hnací silou pro zvýšení produktivity ve stavebnictví.

Lidské myšlení je organizováno na podobném principu: při řešení nových problémů vycházíme z dříve získaných znalostí (obr. 4.4-19, obr. 4.4-20, obr. 4.4-21). Ve strojovém učení tento přístup funguje také - zjednodušením datového modelu a jeho elegantnější úpravou můžeme snížit složitost problému pro algoritmy ML. To zase snižuje potřebu velkého množství dat a snižuje výpočetní náklady.

Od subjektivního hodnocení ke statistické předpovědi

Doba, kdy strategická rozhodnutí závisela na intuici jednotlivých manažerů (obrázek 9.2-4), je minulostí. Ve stále konkurenčnějším a náročnějším ekonomickém prostředí se subjektivní přístup stává příliš riskantním a neefektivním. Podniky, které se nadále spoléhají na osobní názory namísto objektivní analýzy údajů, ztrácejí schopnost rychle reagovat na změny.

Konkurenční prostředí vyžaduje přesnost a opakovatelnost na základě dat, statistických vzorců a vypočitatelné pravděpodobnosti. Rozhodnutí již nemohou být založena na pocitu, musí být založena na korelacích, trendech a prediktivních modelech odvozených z analytiky a strojového učení. Nejde jen o změnu nástrojů - jde o změnu logiky myšlení: od předpokladů k důkazům, od subjektivních pravděpodobností ke statisticky vypočítaným odchylkám, od pocitů k faktům.



Obr. 9.2-4 Éra rozhodování podle HiPPO (názor nejlépe placeného zaměstnance) se s nástupem velkých dat a strojového učení stane minulostí.

Manažeři, kteří se dříve spoléhali pouze na své pocity, budou nevyhnutelně čelit nové realitě: autorita již neurčuje volbu. V centru řízení jsou nyní systémy, které analyzují miliony parametrů a vektorů, identifikují skryté vzorce a navrhují optimální strategie.

Hlavním důvodem, proč se dnes společnosti stále vyhýbají zavádění ML, je jeho netransparentnost. Většina modelů funguje pro manažery jako "černé skříňky", aniž by bylo vysvětleno, jak přesně docházejí ke svým závěrům. To vede k problémům: algoritmy mohou posilovat stereotypy a dokonce vytvářet humorné situace, jako v případě chatbota společnosti Microsoft, který se rychle změnil v toxický komunikační nástroj [154].

V knize Deep Thinking se bývalý mistr světa v šachu Garry Kasparov zamýšlí nad svou porážkou od počítače IBM Big Blue [155]. Tvrdí, že skutečná hodnota umělé inteligence nespočívá v kopírování lidské inteligence, ale v doplnění našich schopností. Umělá inteligence by měla plnit úkoly, kde jsou lidé slabí, zatímco lidé přinášejí kreativitu. Počítače změnily tradiční přístup k analýze šachů. Namísto vytváření fascinujících příběhů o partiích počítačové šachové programy hodnotí každý tah nestranně, pouze na základě jeho skutečné síly nebo slabiny. Kasparov poznamenává, že lidská tendence vnímat události spíše jako souvislé příběhy než jednotlivé akce často vede k nesprávným závěrům - nejen v šachu, ale v životě obecně.

Pokud tedy plánujete používat strojové učení pro predikce a analýzy, je důležité pochopit jeho základní principy - jak fungují algoritmy a jak se zpracovávají data - dříve, než začnete při své práci používat nástroje strojového učení a AI. Nejlepším způsobem, jak začít, jsou praktické zkušenosti.

Jedním z nevhodnějších nástrojů pro úvodní seznámení s tématem strojového učení a predikcí je Jupyter Notebook a oblíbená klasická datová sada Titanic, která názorně představí klíčové metody analýzy dat a tvorby ML modelů.

Titanic dataset: Hello World ve světě analytiky data a big data

Jedním z nejznámějších příkladů využití ML v datové analytice je analýza souboru dat z Titaniku, který se často používá ke studiu pravděpodobnosti přežití cestujících. Učení této tabulky je analogické programu "Hello World" při učení programovacích jazyků.

Při potopení RMS Titanic v roce 1912 zahynulo 1502 z 2224 lidí. Soubor údajů o Titaniku obsahuje nejen informace o tom, zda cestující přežil, ale také atributy, jako jsou: věk, pohlaví, třída letenků a další parametry. Tato datová sada je k dispozici zdarma a lze ji otevřít a analyzovat na různých offline i online platformách.

Odkaz na datovou sadu Titanic:

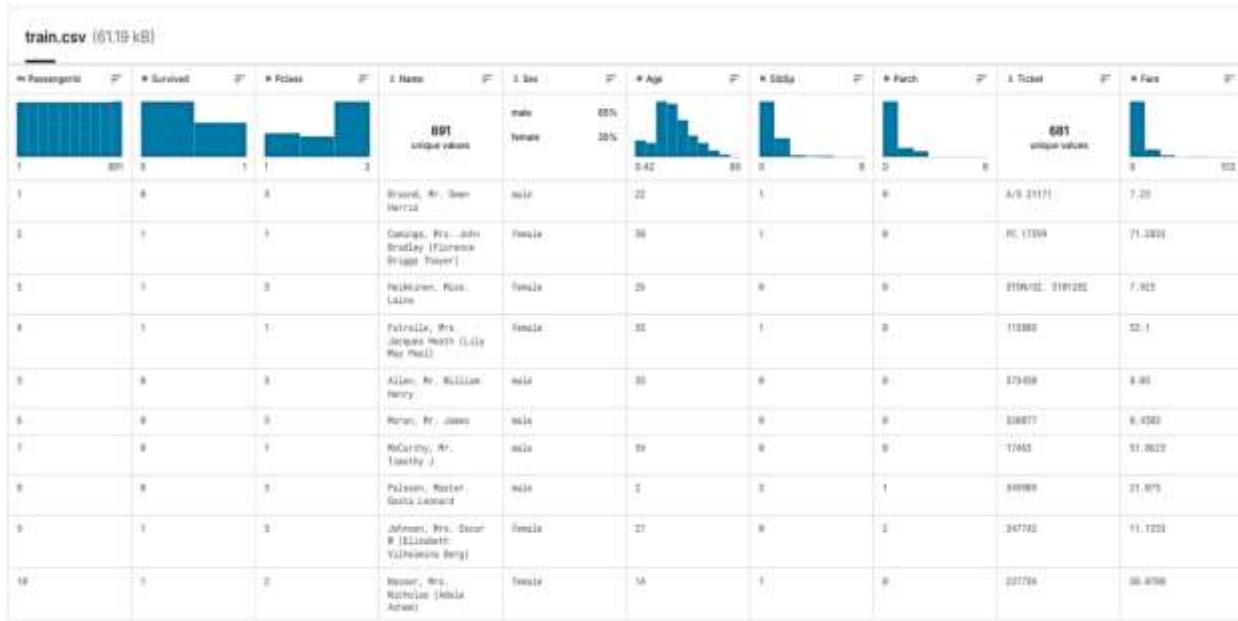
<https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv>

Již dříve v kapitole "IDE s podporou LLM a budoucí změny v programování" jsme se zabývali Jupyter Notebookem - jedním z nejoblíbenějších vývojových prostředí pro analýzu dat a strojové učení. Bezplatnou cloudovou obdobou Jupyter Notebook jsou platformy Kaggle a Google Collab, které umožňují spouštět kód v jazyce Python bez instalace softwaru a poskytují bezplatný přístup k výpočetním zdrojům.

Kaggle je největší soutěžní platforma pro analýzu dat a strojové učení s integrovaným prostředím pro spouštění kódu. K říjnu 2023 má Kaggle více než 15 milionů uživatelů [156] ze 194 zemí.

Stáhněte si a používejte datovou sadu Titanic na platformě Kaggle (obrázek 9.2-5), kde je datová sada (její kopie) uložena a kód Pythonu s předinstalovanými knihovnami lze spouštět přímo v prohlížeči,

aniž byste museli instalovat specializované vývojové prostředí.



Obr. 9.2-5 Statistiky tabulky Titanic - nejoblíbenější trénovací soubor dat pro výuku datové analýzy a strojového učení.

Datová sada Titanic obsahuje údaje o 2224 cestujících na palubě RMS *Titanic* v době jeho ztroskotání v roce 1912. Soubor dat je prezentován jako dvě samostatné tabulky, tréninkový (train.csv) a testovací (test.csv) vzorek, což umožňuje jeho použití jak pro trénování modelů, tak pro vyhodnocení jejich přesnosti na nových datech.

Tréninkový soubor dat obsahuje jak atributy - atributy cestujících (věk, pohlaví, třída jízdenky a další), tak informace o tom, kdo přežil (sloupec s binární hodnotou "Přežil"). Tréninkový dataset (obrázek 9.2-6 - soubor train.csv) se používá k trénování modelu. Testovací dataset (obr. 9.2-7 - soubor test.csv) obsahuje pouze atributy cestujících bez informací o přeživších (bez jediného sloupce "Přežil"). Testovací soubor dat je určen k testování modelu na nových datech a k vyhodnocení jeho přesnosti.

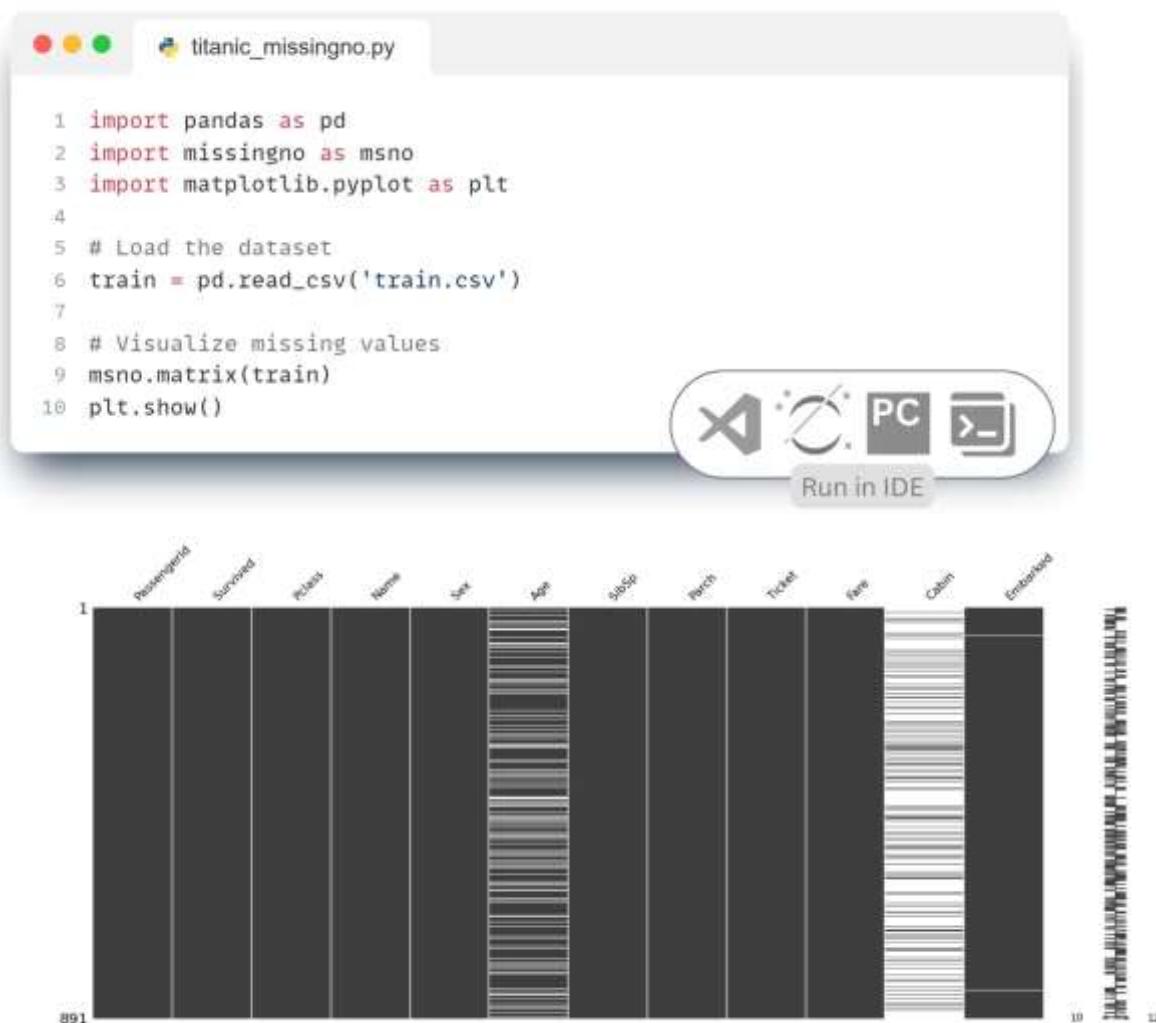
V trénovacích a testovacích souborech dat tak máme téměř identické atributy cestujících. Jediný klíčový rozdíl je v tom, že v testovací datové sadě máme seznam cestujících, kteří nemají sloupec "Survivor" - cílovou proměnnou, kterou se chceme naučit předpovídat pomocí různých matematických algoritmů. A po sestavení modelu budeme moci porovnat výstup našeho modelu se skutečným parametrem "Survivor" z testovacího datasetu, který budeme brát v úvahu při vyhodnocování výsledků.

Hlavní sloupce tabulky, parametry cestujících v trénovací a testovací sadě dat:

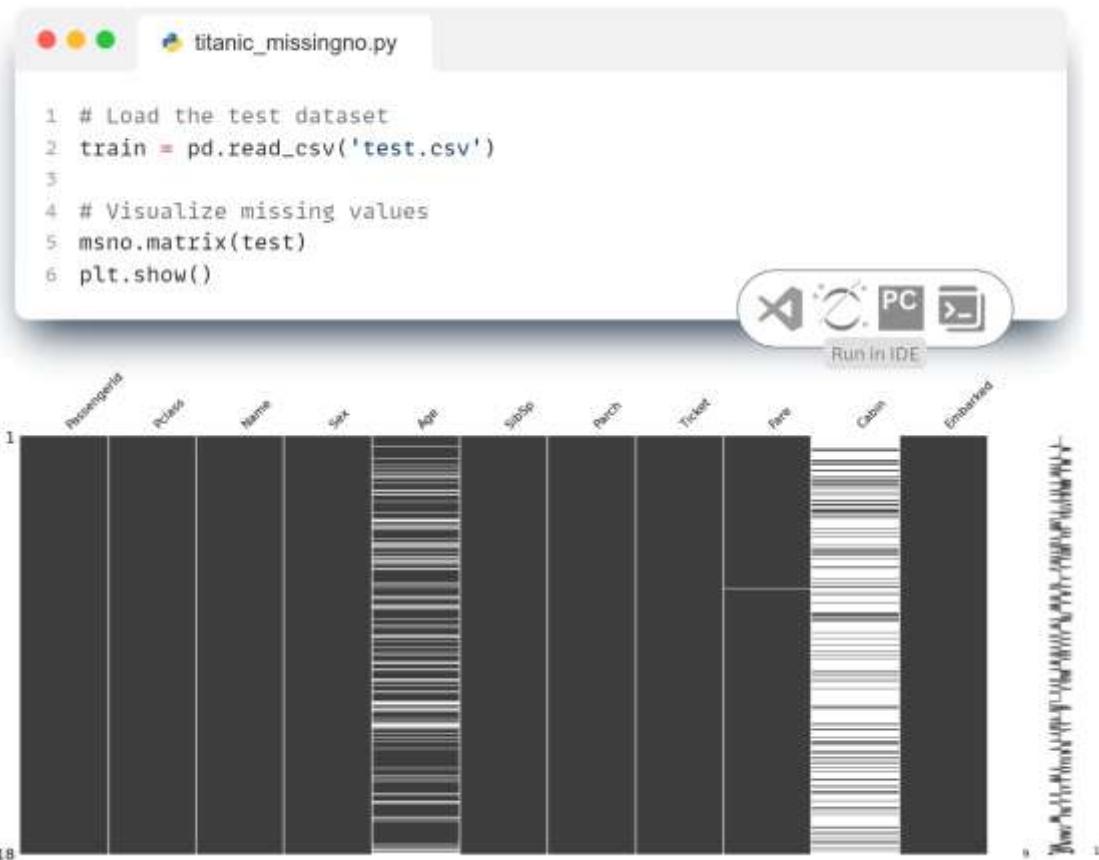
- **PassengerId** - jedinečný identifikátor cestujícího
- **Přežil** - 1, pokud cestující přežil, 0, pokud zemřel (není k dispozici v testovacím souboru)
- **Pclass** - třída jízdenky (1, 2 nebo 3)
- **Jméno** - jméno cestujícího
- **Pohlaví** - pohlaví cestujícího (muž/žena)

- Věk
- **SibSp** - počet sourozenců nebo manželů na palubě lodi
- **Parch** - počet rodičů nebo dětí na lodi
- **Vstupenka** - číslo vstupenky
- **Jízdné** - cena jízdenky
- Kabina - číslo kabiny (mnoho údajů chybí)
- **Embarked** je přístav nalodění (C = Cherbourg, Q = Queenstown, S = Southampton).

Pro vizualizaci chybějících dat v obou tabulkách můžete použít knihovnu missingno (obr. 9.2-6, obr. 9.2-7), která zobrazuje chybějící hodnoty ve formě histogramu, kde bílá pole označují chybějící data. Tato vizualizace umožňuje rychlé posouzení kvality dat před jejich zpracováním.



Obr. 9.2-6 Několik řádků kódu slouží k vizualizaci chybějících dat v trénovací sadě dat Titanic, kde je klíčovým parametrem pro trénování parametr "Survived".



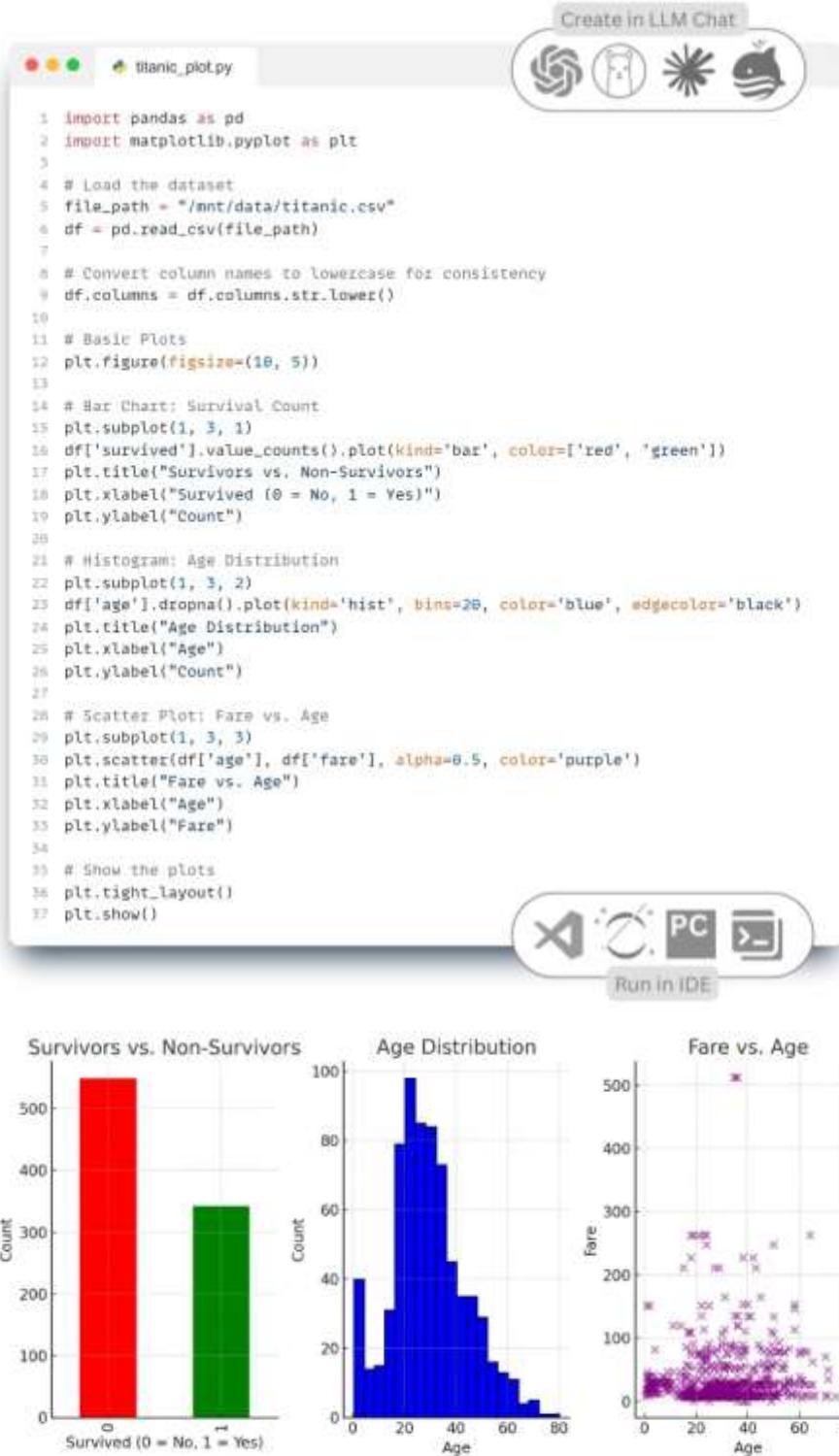
Obrázek 9.2-7 Vizualizace chybějících údajů v testovacím souboru dat Titanic, který obsahuje pouze charakteristiky cestujících bez informací.

Před formulováním hypotéz a vytvářením předpovědí na základě souboru dat pomáhají vizuální analýzy identifikovat klíčové vzorce v datech, posoudit jejich kvalitu a určit možné závislosti. Existuje mnoho vizualizačních technik, které vám mohou pomoci lépe porozumět souboru dat Titanic. K posouzení kvality informací a pochopení dat můžete použít distribuční grafy pro analýzu věkových skupin cestujících, grafy přežití podle pohlaví a třídy a matice chybějících údajů.

- 💡 Požádejme LLM, aby nám pomohl vizualizovat data ze sady dat Titanic, a to zasláním následujícího textového požadavku libovolnému modelu LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN nebo jinému):

Ukažte prosím několik jednoduchých grafů pro soubor dat Titanic. Sami si stáhněte datovou sadu a ukažte ↴

- 💡 Odpovědi LLM ve formě připraveného kódu a grafů vizualizujících parametry datové sady



Obrázek 9.2-8 LLM vám pomůže získat okamžitou vizualizaci dat vaší datové sady.

Vizualizace dat je důležitým krokem k přípravě datové sady pro následnou konstrukci modelu

strojového učení, k němuž lze přistoupit pouze na základě porozumění datům.

Strojové učení v akci: od pasažérů Titaniku po řízení projektů

Hlavní hypotézou použitou pro zkoumání rámce strojového učení na základě souboru dat z Titaniku je, že určité skupiny cestujících měly větší šanci na přežití.

Malá tabulka pasažérů Titaniku se stala populární po celém světě a miliony lidí ji používají k trénování, experimentování a testování modelů, aby zjistili, které algoritmy a hypotézy vytvoří nejpřesnější model předpovědi přežití na základě tréninkového souboru dat pro pasažéry Titaniku.

Přitažlivost souboru dat Titanic spočívá v jeho kompaktnosti: s několika stovkami řádků a dvanácti sloupci (obr. 9.2-6) poskytuje dostatek příležitostí k analýze. Datová sada je, poměrně jednoduše řečeno, klasickým příkladem řešení binární klasifikace, kde je cíl problému - přežití - vyjádřen v příhodném formátu 0 nebo 1.

John Wheeler ve filmu "It from Bit" [7] tvrdí, že vesmír je založen na binárních volbách. Podobně i podnik řízený lidmi složenými z molekul je ve skutečnosti postaven na řadě binárních binárních voleb.

Kromě toho jsou data založena na skutečné historické události, což je na rozdíl od uměle vytvořených příkladů činí cennými pro výzkum. Jen na platformě Kaggle, která je jednou z největších databází Data Pipeline a ETL, se výzev založených na datové sadě Titanic zúčastnilo 1 355 998 lidí, kteří vytvořili 53 963 unikátních řešení Data Pipeline [157] (obrázek 9.2-9).

Zdá se to neuvěřitelné, ale pouhých 1000 řádků dat o pasažérech Titaniku s 12 parametry se stalo polem pro miliony hypotéz, logických řetězců a jedinečných datových potrubí. Z malého souboru dat se rodí nekonečné množství poznatků, hypotéz a interpretací - od jednoduchých modelů přežití až po složité soubory, které zohledňují skryté zákonitosti a komplexní labyrinty uvažování.

Machine Learning from Disaster

[Submit Prediction](#)

Data Code Models Discussion Leaderboard Rules

Titanic Tutorial 16916
Updated 3y ago Gold ***
29858 comments · Titanic - Machine Learning from Disaster

Titanic competition w/ TensorFlow Decision Forests 1098
Updated 2y ago Gold ***
Score: 0.80143 · 318 comments · Titanic - Machine Learning from Disaster

Titanic Data Science Solutions 10723
Updated 6y ago Gold ***
2590 comments · Titanic - Machine Learning from Disaster

Exploring Survival on the Titanic 3968
Updated 7y ago Gold ***
Score: 0.80382 · 1072 comments · Titanic - Machine Learning from Disaster

Obr. 9.2-9 Prvních pět řešení z celkového počtu 53 963 hotových a otevřených řešení Pipeline. Jen na serveru Kaggle [157] se tento problém pokusilo vyřešit již téměř 1,5 milionu lidí.

Jestliže i takto malá tabulka dokáže vygenerovat miliony unikátních řešení (obr. 9.2-9), co můžeme říci o skutečných průmyslových konstrukčních souborech dat, kde se parametry měří v desítkách tisíc?

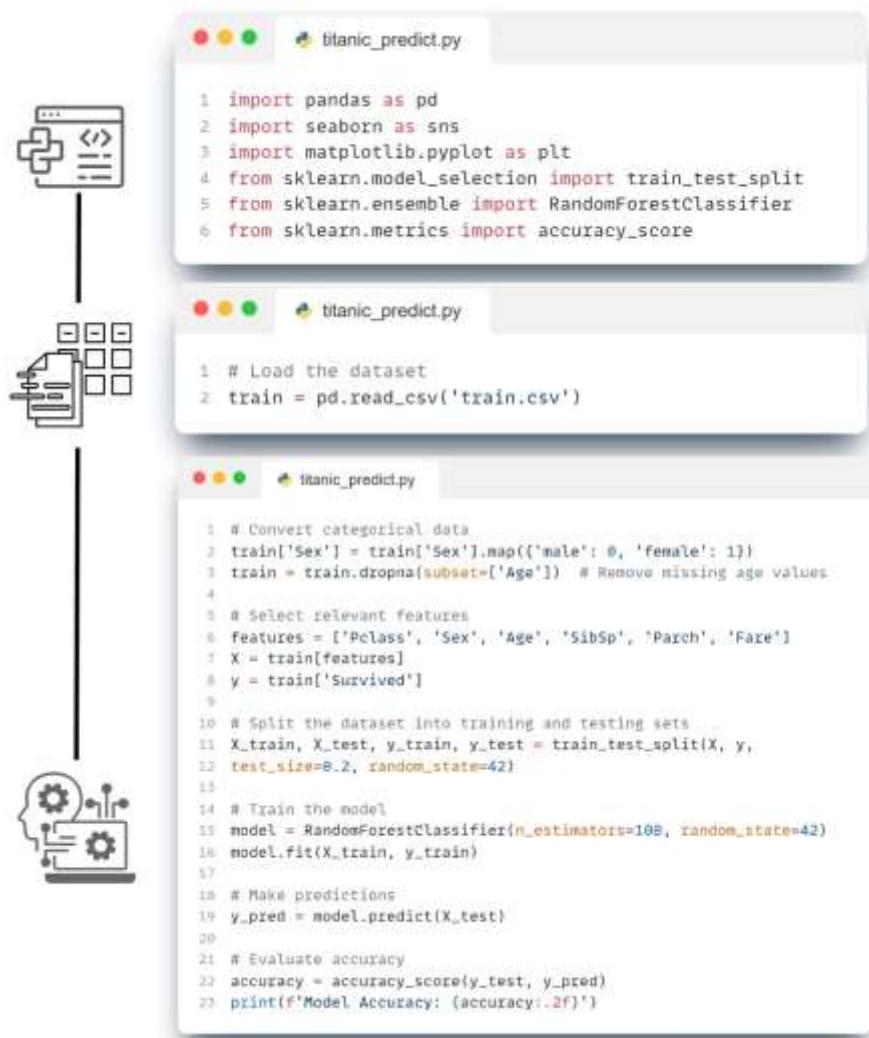
Standardní CAD -návrh relativně malé budovy obsahuje desítky tisíc entit s tisíci parametry - od geometrických charakteristik po nákladové a časové atributy. Představte si, kolik potenciálních poznatků, vztahů, předpovědí a manažerských hypotéz se skrývá v datech ze všech projektů vaší společnosti shromázděných v průběhu let. Historická projektová data nejsou jen archivem - je to živá paměť organizace, její digitální stopa, kterou lze analyzovat a vytvořit tak velké množství jedinečných hypotéz.

A co je nejdůležitější, nemusíte čekat, až se komunita Kaggle začne zajímat o vaši společnost nebo vaše data. Můžete začít pracovat s tím, co máte již dnes: spouštět analýzy na vlastních datech, trénovat modely na vlastních datech, identifikovat opakování, anomálie a vzory. Tam, kde dříve byly zapotřebí roky experimentů a drahé konzultace, dnes stačí iniciativa, titul LLM, otevřený přístup k datům a ochota učit se.

- ➲ Abychom mohli sestavit algoritmus strojového učení, který bude předpovídat míru přežití cestujících na základě trénovacího souboru dat o cestujících train.csv, požádejme LLM, aby tento problém vyřešil za nás:

Na základě trénovacího souboru dat o pasažérech Titaniku vytvořte model strojového učení pro předpovídání schopnosti přežít. ↗

2 Odpověď společnosti LLM :



```

titanic_predict.py
1 import pandas as pd
2 import seaborn as sns
3 import matplotlib.pyplot as plt
4 from sklearn.model_selection import train_test_split
5 from sklearn.ensemble import RandomForestClassifier
6 from sklearn.metrics import accuracy_score

# Load the dataset
train = pd.read_csv('train.csv')

# Convert categorical data
train['Sex'] = train['Sex'].map({'male': 0, 'female': 1})
train = train.dropna(subset=['Age']) # Remove missing age values

# Select relevant features
features = ['Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Fare']
X = train[features]
y = train['Survived']

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Train the model
model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)

# Evaluate accuracy
accuracy = accuracy_score(y_test, y_pred)
print(f'Model Accuracy: {accuracy:.2f}')

```

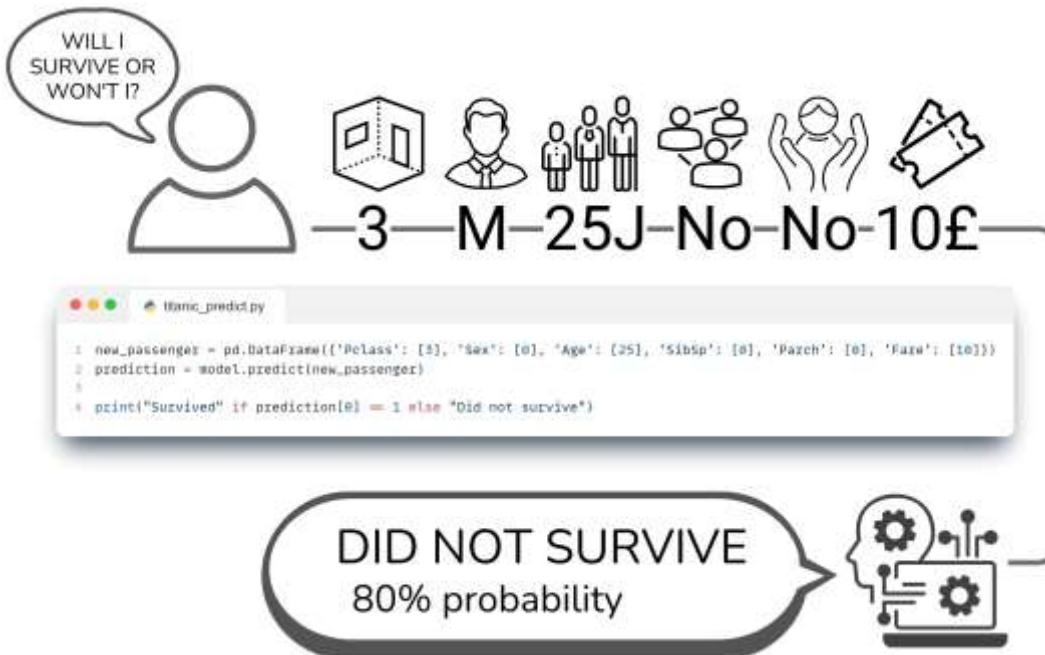
Obr. 9.2-10 LLM sestavil předpověď přeživších z Titaniku pomocí algoritmu strojového učení Random Forest.

Výsledný kód z LLM (obr. 9.2-10) načte data o pasažérech Titaniku, vyčistí je, převeďe kategoriální proměnné (např. pohlaví do číselného formátu) a vycvičí model pomocí algoritmu RandomForestClassifier, aby předpovíděl, zda pasažér přežil, či nikoli (o populárních algoritmech si povíme více v následujících kapitolách).

Kód rozdělí trénovací data na trénovací a testovací množiny (na webových stránkách Kaggle jsou již vytvořeny hotové soubory test.csv (obr. 9.2-7) a train.csv (obr. 9.2-6) pro trénování, poté se model natrénuje na trénovacích datech a otestuje na testovacích datech, aby se zjistilo, jak dobrý je konkrétní predikční model. Po tréninku se do modelu vloží testovací data ze souboru test.csv (se skutečnými údaji o těch, kteří přežili nebo nepřežili) a model předpovídá, kdo přežil a kdo ne. V našem případě je přesnost modelu strojového učení, který jsme získali, přibližně 80%, což ukazuje, že poměrně dobře vystihuje zákonitosti.

Strojové učení lze přirovnat k dítěti, které se snaží vměstnat obdélníkovou kostku do kulaté díry. V počátečních fázích algoritmus zkouší mnoho přístupů, přičemž narází na chyby a nesrovnalosti. Tento proces se může zdát neefektivní, ale poskytuje důležité učení: analýzou každé chyby model zlepšuje své předpovědi a činí stále přesnější rozhodnutí.

Nyní lze tento model (obr. 9.2-10) použít k předpovědi míry přežití nových cestujících a například pokud jej nakrmíte informacemi o cestujících pomocí funkce model.predict s parametry: "muž", "3. třída", "25 let", "bez příbuzných na palubě", model vytvoří předpověď - že cestující s 80% pravděpodobností nepřežije katastrofu, pokud byl na Titaniku v roce 1912 (obr. 9.2-11).



Obrázek 9.2-11 Model, který jsme vytvořili výše, nyní dokáže s 80% pravděpodobností předpovědět, zda některý z nových pasažérů Titaniku přežije.

Model předpovědi přežití pasažérů Titaniku ilustruje mnohem širší koncept: tisíce profesionálů ve stavebnictví dělají každý den podobná "dvojí" rozhodnutí - o životě nebo smrti rozhodnutí, projektu, odhadu, nástroje, zisku nebo ztrátě, bezpečnosti nebo riziku. Stejně jako v příkladu Titaniku, kde výsledek závisel na faktorech (pohlaví, věk, třída), i ve stavebnictví je každý aspekt rozhodnutí ovlivněn

mnoha vlastními faktory a proměnnými (sloupci tabulek): náklady na materiál, kvalifikace pracovníků, termíny, počasí, logistika, technická rizika, připomínky a statisíce dalších parametrů.

Ve stavebnictví se strojové učení řídí stejnými principy jako v jiných oborech: modely se trénují na historických datech - z projektů, zakázek, odhadů - s cílem testovat různé hypotézy a najít nejfektivnější řešení. Tento proces se podobá učení dítěte metodou pokusů a omylů: s každým dalším cyklem se modely přizpůsobují a stávají se přesnějšími.

Využití nashromážděných dat otevírá nové obzory pro stavebnictví. Namísto časově náročných manuálních výpočtů lze vycvičit modely, které dokáží s vysokou přesností předpovídat klíčové charakteristiky budoucích projektů. Prediktivní analytika tak mění stavebnictví v prostor, kde lze nejen plánovat, ale také s jistotou předvídat vývoj.

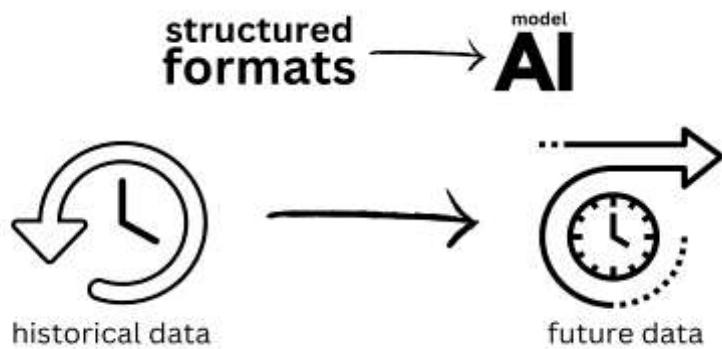
Předpovědi a prognózy na základě historických dat

Údaje shromážděné o projektech společnosti otevírají možnost sestavit modely schopné předpovídat nákladové a časové charakteristiky budoucích, dosud nerealizovaných objektů - bez časově náročných ručních výpočtů a porovnávání. To umožňuje výrazně urychlit a zjednodušit procesy oceňování, přičemž se nespolehlá na subjektivní předpoklady, ale na spolehlivé matematické předpovědi.

Dříve, ve čtvrté části knihy, jsme se podrobně zabývali tradičními metodami odhadu nákladů projektu, včetně metody založené na zdrojích, a zmínili jsme se také o parametrických a expertních přístupech. Tyto metody jsou stále aktuální, ale v moderní praxi se začínají obohacovat o nástroje statistické analýzy a strojového učení, které mohou výrazně zlepšit přesnost a reprodukovatelnost odhadů.

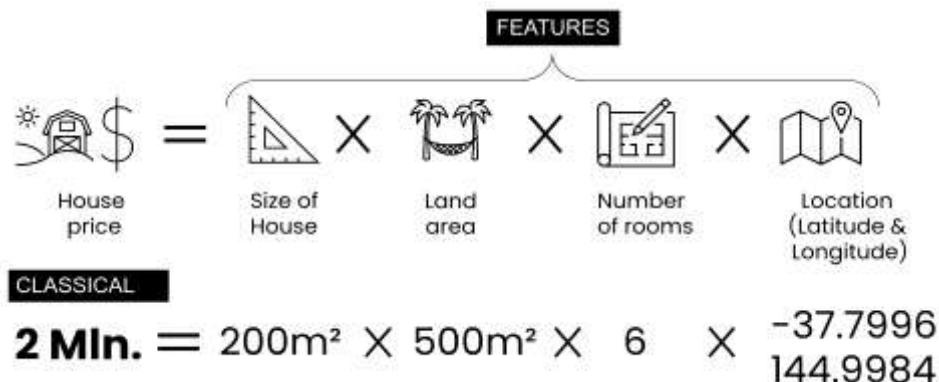
Procesy manuálního a poloautomatického výpočtu cen a časových atributů budou v budoucnosti doplněny o názory a předpovědi ML modelů schopných analyzovat historická data, nacházet skryté vzorce a navrhovat informovaná rozhodnutí. Nová data a scénáře budou generovány automaticky z již dostupných informací - podobně jako jazykové modely (LLM) vytvářejí texty, obrázky a kódy na základě dat shromážděných v průběhu let z otevřených zdrojů [158].

Stejně jako se dnes lidé při posuzování budoucích událostí spoléhají na zkušenosti, intuici a interní statistiky, bude v příštích letech budoucnost stavebních projektů stále více určovat kombinace nashromážděných znalostí a matematických modelů strojového učení.



Obr. 9.2-12 Kvalitativní a strukturovaná historická firemní data jsou materiálem, na kterém jsou postaveny modely strojového učení a předpovědi.

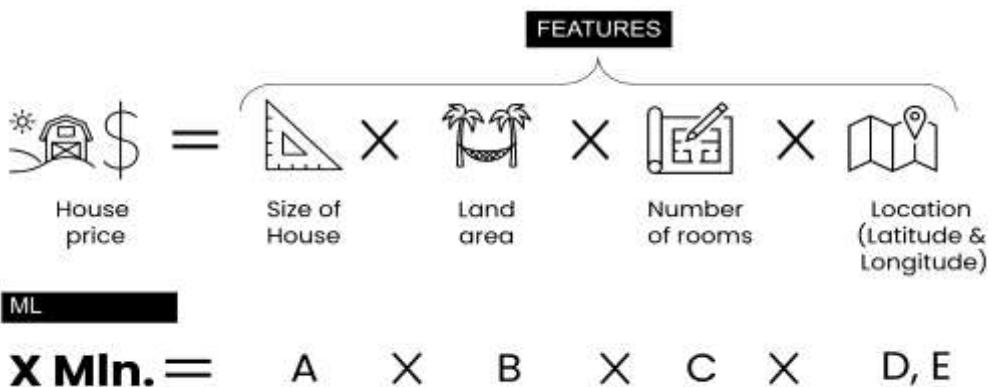
Vezměme si jednoduchý příklad: předpověď ceny domu na základě jeho plochy, velikosti pozemku, počtu místností a zeměpisné polohy. Jedním z přístupů je sestavení klasického modelu, který analyzuje tyto parametry a vypočítá očekávanou cenu (obrázek 9.2-13). Tento přístup vyžaduje přesný a předem známý vzorec, což je v reálné praxi prakticky nemožné.



Obr. 9.2-13 K odhadu hodnoty domu lze použít klasický algoritmus s pevně daným vzorcem, který je třeba najít.

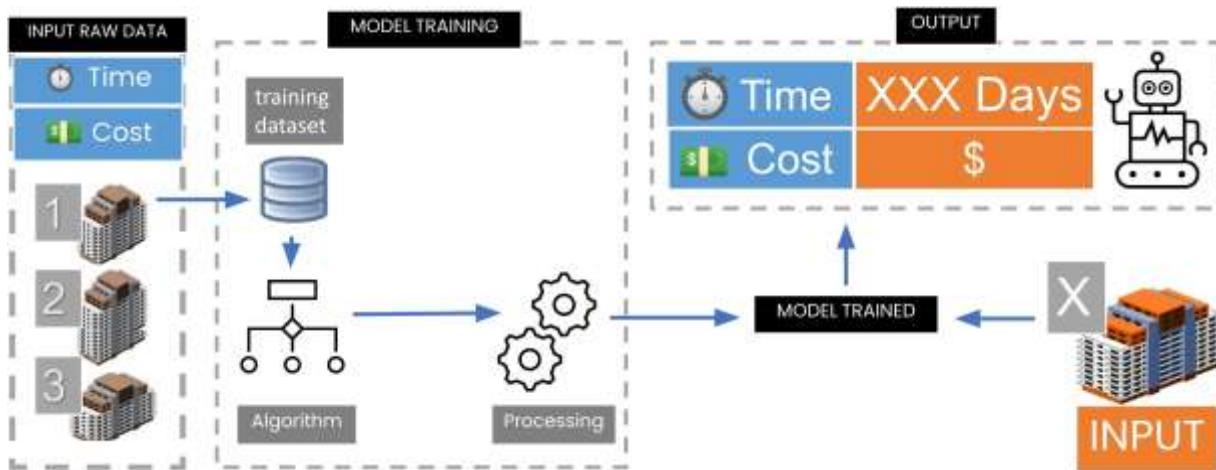
Strojové učení eliminuje ruční vyhledávání vzorců a nahrazuje je vyškolenými algoritmy, které nezávisle identifikují závislosti, jež jsou mnohonásobně přesnější než jakékoli předem stanovené rovnice. Případně vytvořme algoritmus strojového učení, který vytvoří model na základě předchozího pochopení problému a historických dat, která mohou být neúplná (obrázek 9.2-14).

Na příkladu cen umožňuje strojové učení vytvářet různé typy matematických modelů, které nevyžadují znalost přesného mechanismu tvorby nákladů. Model se "učí" z údajů o předchozích projektech a přizpůsobuje se skutečným zákonitostem mezi parametry staveb, jejich náklady a termíny.



Obr. 9.2-14 Na rozdíl od klasického odhadu založeného na vzorcích je algoritmus strojového učení trénován na historických datech.

V kontextu strojového učení pod dohledem obsahuje každý projekt v souboru trénovacích dat jak vstupní atributy (např. náklady a čas podobných staveb), tak očekávané výstupní hodnoty (např. náklady nebo čas). Podobná datová sada se používá k vytvoření a přizpůsobení modelu strojového učení (obrázek 9.2-15). Čím větší je datová sada a čím vyšší je kvalita dat v ní obsažených, tím přesnější bude model a tím přesnější budou výsledky predikce.



Obrázek 9.2-15 ML model natrénovaný na datech o nákladech a harmonogramu z minulých projektů určí náklady a harmonogram nového projektu s určitou pravděpodobností.

Jakmile je model vytvořen a vycvičen k odhadu výstavby nového projektu, stačí modelu zadat nové atributy nového projektu a model poskytne odhadované výsledky na základě dříve naučených vzorů s určitou pravděpodobností.

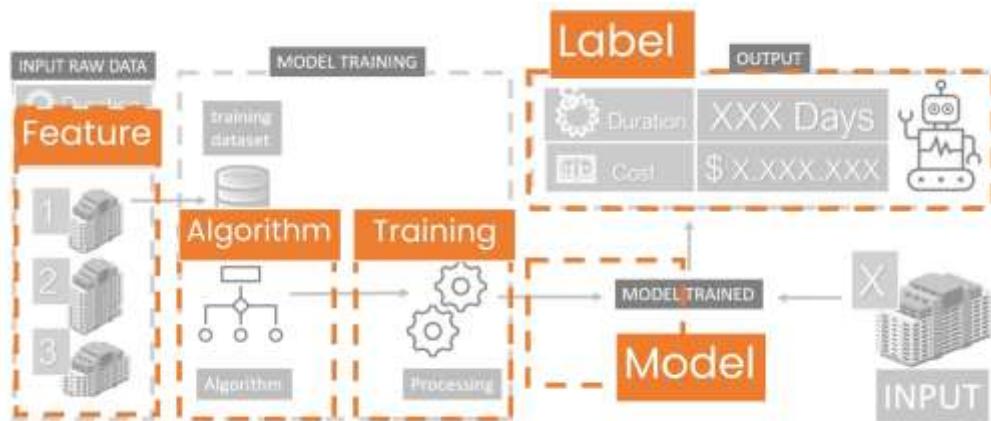
Klíčové koncepty strojového učení

Strojové učení není magie, je to jen matematika, data a hledání vzorců. Nemá žádnou skutečnou inteligenci, ale je to program vycvičený na datech, který rozpoznává vzory a rozhoduje bez neustálého

zapojení člověka.

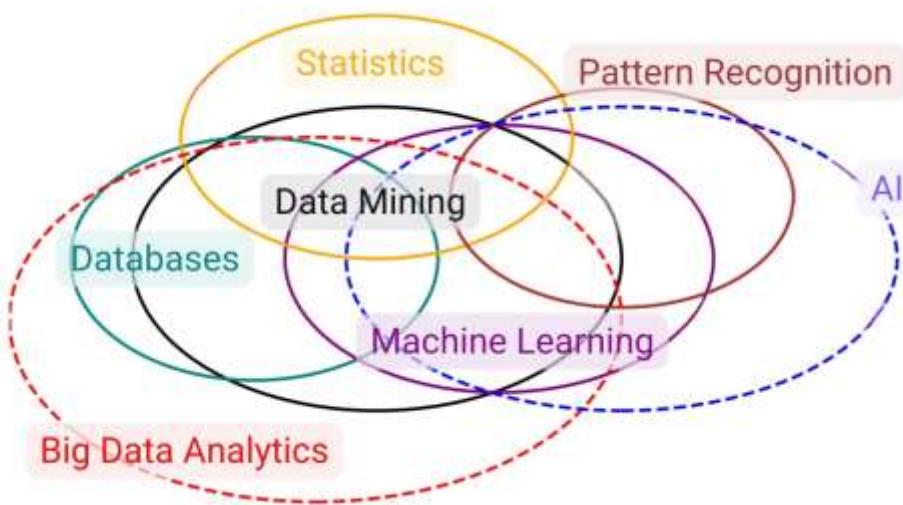
Strojové učení používá k popisu své struktury několik klíčových pojmu (obrázek 9.2-16):

- **Štítky** jsou cílové proměnné nebo atributy (parametr "Survivor" v datové sadě Titanic), které by měl model předpovídat. Příklad: stavební náklady (např. v dolarech), doba trvání stavebních prací (např. v měsících).
- **Funkce** jsou nezávislé proměnné nebo atributy, které slouží jako vstupy do modelu. V prognostickém modelu se používají k předpovídání značek. Příklady: velikost pozemku (v metrech čtverečních), počet podlaží budovy, celková podlahová plocha budovy (v metrech čtverečních), zeměpisná poloha (zeměpisná šířka a délka), typ materiálů použitých při stavbě. Počet charakteristik také určuje dimenzionalitu dat.
- **Model** je soubor různých hypotéz, z nichž jedna approximuje cílovou funkci, kterou je třeba předpovědět nebo approximovat. Příklad: model strojového učení, který využívá techniky regresní analýzy k předpovídání nákladů a termínů výstavby.
- **Algoritmus učení Algoritmus učení** je proces hledání nejlepší hypotézy v modelu, která přesně odpovídá cílové funkci, pomocí sady trénovacích dat. Příklad: Algoritmus lineární regrese, KNN nebo náhodného lesa, který analyzuje údaje o nákladech a době výstavby s cílem identifikovat vztahy a vzorce.
- **Trénování** - během procesu trénování algoritmus analyzuje trénovací data s cílem najít vzory, které odpovídají vztahu mezi vstupními atributy a cílovými štítky. Výsledkem tohoto procesu je natrénovaný model strojového učení, připravený k predikci. Příklad: Proces, při kterém algoritmus analyzuje historická data o stavbě (náklady, čas, charakteristiky objektu) za účelem vytvoření predikčního modelu.



Obrázek 9.2-16 ML používá štítky a atributy k vytváření modelů, které jsou trénovány na datech pomocí algoritmů pro předpovídání výsledků.

Strojové učení neexistuje izolovaně, ale je součástí širšího ekosystému analytických disciplín, včetně statistiky, databází, dolování dat, rozpoznávání vzorů, analýzy velkých dat a umělé inteligence. Obrázek 9.2-17 ukazuje, jak se tyto obory vzájemně překrývají a doplňují, a poskytují tak komplexní rámec pro moderní rozhodovací a automatizační systémy.

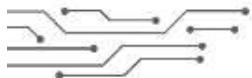


Obr. 9.2-17 Vztah mezi různými oblastmi analýzy dat: statistika, strojové učení, umělá inteligence, big data, rozpoznávání vzorů a dolování dat.

Hlavním cílem strojového učení je vybavit počítače schopností automaticky se učit znalostem bez lidského zásahu nebo pomoci a podle toho upravovat své činnosti [159].

V budoucnu tedy bude úloha člověka spočívat pouze v tom, že poskytne stroji kognitivní schopnosti - nastaví podmínky, váhy a parametry a o zbytek se postará model strojového učení.

V příští kapitole se budeme zabývat konkrétními příklady použití algoritmů. Na reálných tabulkách a zjednodušených modelech si ukážeme, jak se prognóza sestavuje krok za krokem.



KAPITOLA 9.3.

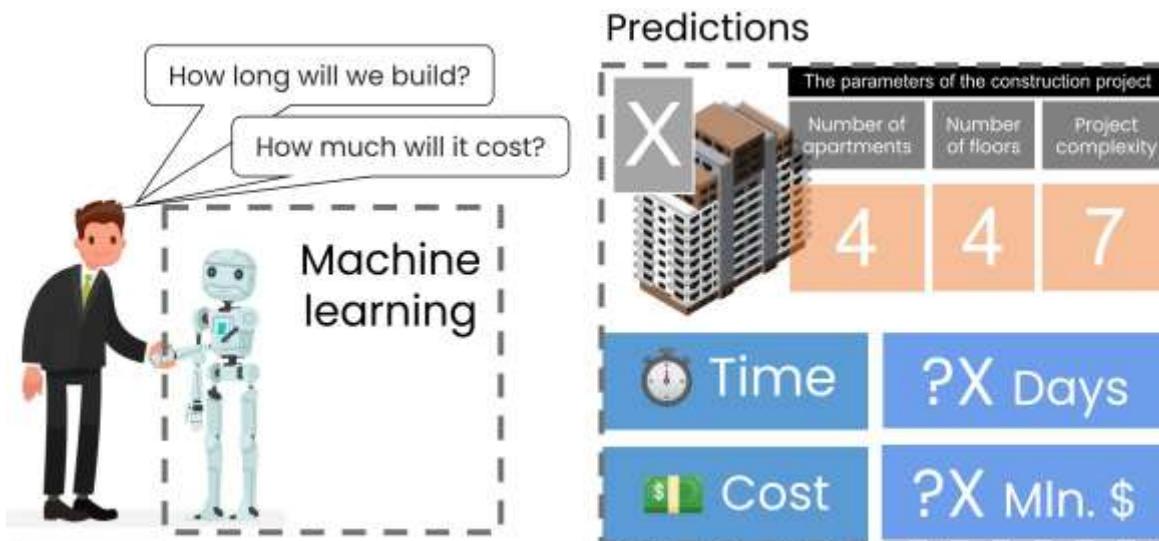
PŘEDPOVÍDÁNÍ NÁKLADŮ A TERMÍNŮ POMOCÍ STROJOVÉHO UČENÍ

Příklad použití strojového učení pro zjištění nákladů a časového harmonogramu projektu

Odhad času a nákladů na stavbu je jedním z klíčových procesů v činnosti stavební firmy. Tradičně tyto odhady provádějí odborníci na základě zkušeností, referenčních knih a regulačních databází. S digitální transformací a rostoucí dostupností dat je však nyní možné využít modely strojového učení (ML) ke zvýšení přesnosti a automatizaci takových odhadů.

Zavedení strojového učení do procesu výpočtu nákladů a termínů výstavby umožňuje nejen efektivnější plánování, ale stává se také východiskem pro integraci inteligentních modelů do dalších podnikových procesů - od řízení rizik až po optimalizaci logistiky a zásobování.

Důležité je umět rychle určit, jak dlouho bude stavba projektu trvat a jaké budou její celkové náklady. Tyto otázky týkající se času a nákladů projektu jsou tradičně v popředí zájmu klientů i stavebních firem již od vzniku stavebnictví.



Obr. 9.3-1 Ve stavebních projektech jsou rychlosť a kvalita odhadu času výstavby a nákladů klíčovými faktory úspěchu.

V následujícím příkladu budou klíčové údaje z minulých projektů extrahovány a použity k vývoji modelu strojového učení, který nám umožní použít model k odhadu nákladů a načasování nových stavebních projektů s novými parametry (obrázek 9.3-1).

Uvažujme tři projekty se třemi klíčovými atributy: počtem bytů (kde 100 bytů odpovídá číslu 10 pro snadnější vizualizaci), počtem podlaží a podmíněnou mírou složitosti stavby na stupnici od 1 do 10, kde 10 je nejvyšší hodnocení složitosti. Ve strojovém učení se proces převodu a zjednodušení hodnot,

jako je 100 na 10 nebo 50 na 5, nazývá "normalizace".

Normalizace ve strojovém učení je proces, při kterém se různá číselná data uvedou do společného měřítka, aby se usnadnilo jejich zpracování a analýza. Tento proces je důležitý zejména tehdy, když mají data různá měřítka a měrné jednotky.

Předpokládejme, že první projekt (obrázek 9.3-2) měl 50 bytů (po normalizaci 5), 7 podlaží a skóre složitosti 2, což znamenalo relativně jednoduchou stavbu. Druhý projekt měl již 80 bytů, 9 podlaží a relativně složitou konstrukci. Za těchto podmínek trvala výstavba prvního a druhého bytového domu 270 a 330 dní a celkové náklady projektu činily 4,5 milionu dolarů, resp. 5,8 milionu dolarů.

Construction project	The parameters of the construction project			The key parameters of the project	
	Number of apartment	Number of floors	Project complexity	Time	Cost
1	50	7	2	270	\$ 4.502.000
2	80	9	6	330	\$ 5.750.000
3	30	5	3	230	\$ 3.262.000
X	40	4	7	?X	\$?X. XXX.XXX

Obrázek 9.3-2 Příklad souboru minulých projektů, který bude použit pro odhad času a nákladů budoucího projektu X.

Při sestavování modelu strojového učení pro taková data je hlavním úkolem identifikovat kritické atributy (nebo značky) pro predikci, v tomto případě dobu výstavby a náklady. S malým souborem dat využijeme informace o předchozích stavebních projektech k plánování nových: pomocí algoritmů strojového učení, musíme předpovědět náklady a dobu výstavby nového projektu X na základě daných atributů nového projektu, jako je 40 bytů, 4 podlaží a relativně vysoká složitost projektu 7 (obr. 9.3-2). V reálném prostředí může být počet vstupních parametrů mnohem větší, od několika desítek až po stovky faktorů. Mohou to být například: typ stavebních materiálů, klimatické pásmo, kvalifikační úroveň dodavatelů, dostupnost inženýrských sítí, typ založení, roční období zahájení prací, připomínky mistrů atd.

Abychom mohli vytvořit prediktivní model strojového učení, musíme zvolit algoritmus pro jeho vytvoření. Algoritmus ve strojovém učení je jako matematický recept, který učí počítač, jak provádět předpovědi (míchat ve správném pořadí parametry) nebo se rozhodovat na základě dat.

K analýze dat o minulých stavebních projektech a předpovědi času a nákladů budoucích projektů (obrázek 9.3-2) lze použít jeden z populárních algoritmů strojového učení:

- **Lineární regrese (Lineární regrese):** tento algoritmus se snaží najít přímý vztah mezi atributy, například mezi počtem podlaží a stavebními náklady. Cílem algoritmu je najít lineární rovnici, která tento vztah nejlépe popisuje a umožňuje provádět předpovědi.
- **Algoritmus k-nejbližší sousedé (k-NN):** tento algoritmus porovnává nový projekt s minulými projekty, které byly podobné svou velikostí nebo složitostí. K-NN klasifikuje data na základě toho, které z k (počtu) trénovacích příkladů jsou mu nejbliže. V kontextu regrese je výsledkem průměr nebo medián k nejbližších sousedů.
- **Rozhodovací stromy:** je prediktivní model, který rozděluje data do podmnožin na základě různých podmínek pomocí stromové struktury. Každý uzel stromu představuje podmínu nebo otázku vedoucí k dalšímu rozdělení dat a každý list představuje konečnou předpověď nebo výsledek. Algoritmus rozděluje data do menších skupin na základě různých charakteristik, například nejprve podle počtu příběhů, pak podle složitosti atd. a na základě toho vytváří předpověď.

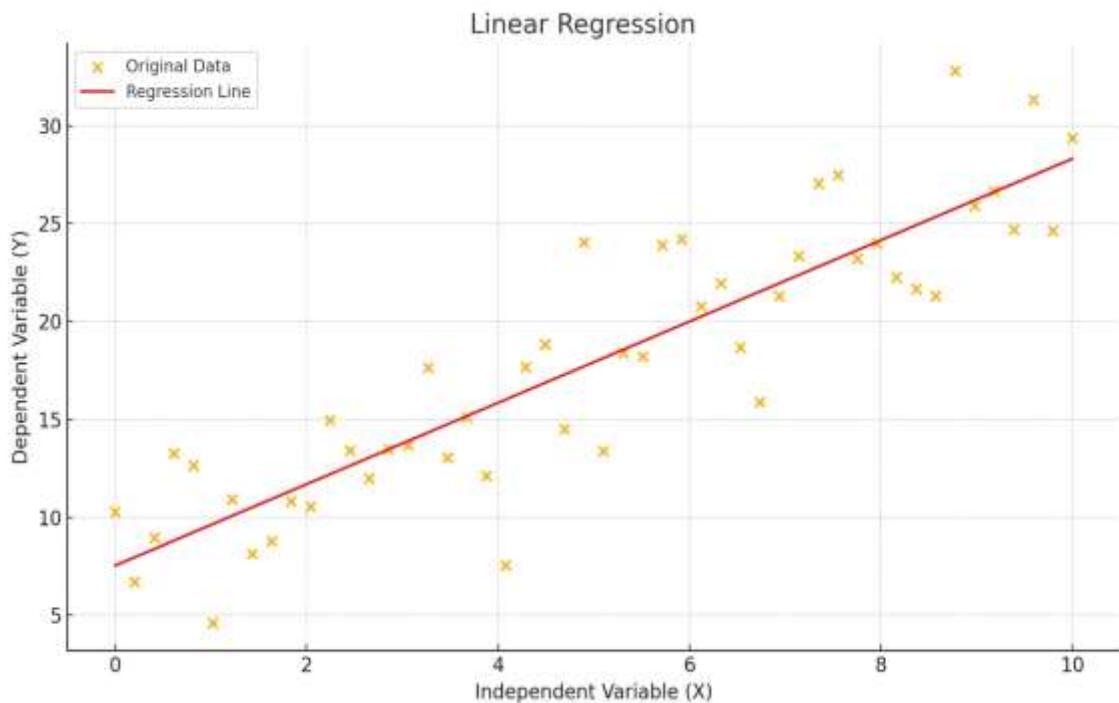
Podívejme se na algoritmy strojového učení pro odhad nákladů na nový projekt na příkladu dvou populárních algoritmů: lineární regrese a algoritmu K-nejbližších sousedů.

Předpověď nákladů a času projektu pomocí lineární regrese

Lineární regrese je základní algoritmus analýzy dat, který předpovídá hodnotu proměnné na základě lineárního vztahu s jednou nebo více jinými proměnnými. Tento model předpokládá, že mezi závislou proměnnou a jednou nebo více nezávislými proměnnými existuje přímý lineární vztah, a cílem algoritmu je tento vztah najít.

Jednoduchost a přehlednost lineární regrese z ní učinila oblíbený nástroj v různých oblastech. Při práci s jednou proměnnou je lineární regrese o nalezení nejlépe odpovídající přímky procházející datovými body.

Lineární regrese najde nejlepší přímku (červená čára), která approximuje závislost mezi vstupní proměnnou X a výstupní proměnnou Y. Tato přímka umožňuje předpovídat hodnoty Y pro nové hodnoty X na základě zjištěné lineární závislosti (obrázek 9.3-3).



Obr. 9.3-3 Principem lineární regrese je nalezení nejlepší přímky procházející trénovacími hodnotami.

Tato přímka je znázorněna rovnicí, v níž zadáním hodnoty nezávislé proměnné (X) získáme předpokládanou hodnotu závislé proměnné (Y). Tento postup umožnuje efektivně předpovídat Y ze známých hodnot X pomocí lineárního vztahu mezi nimi. Příklad nalezení takové statisticky zprůměrované přímky lze vidět v hodnocení dat o stavebních povoleních v San Franciscu (obrázek 9.1-7), kde byla pro různé typy zařízení vypočtena inflace pomocí lineární regrese.

Načtěme tabulku dat projektu (obrázek 9.3-2 z předchozí kapitoly) přímo do LLM a požádejme jej, aby pro nás sestavil jednoduchý model strojového učení.

- 👉 Pošlete textovou žádost do chatu LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

Je třeba ukázat konstrukci jednoduchého modelu strojového učení pro předpověď nákladů a času na realizaci nového projektu X (obr. 9.3-2 jako přiložený obrázek). ↴

■ Odpověď společnosti LLM :

```

1  from sklearn.linear_model import LinearRegression
2  import numpy as np
3
4  # Data from the image
5  # Features: Number of apartments, number of floors, project complexity
6  X = np.array([
7      [5, 7, 2],
8      [8, 9, 6],
9      [3, 5, 3]
10 ])
11
12 # Targets: Time (days), Cost ($)
13 y_time = np.array([270, 330, 230])
14 y_cost = np.array([4502000, 5750000, 3262000])
15
16 # Creating linear regression models for time and cost
17 model_time = LinearRegression()
18 model_cost = LinearRegression()
19
20 # Training the models
21 model_time.fit(X, y_time)
22 model_cost.fit(X, y_cost)
23
24 # The new project's features
25 X_new = np.array([[4, 4, 7]])
26
27 # Predicting the time and cost for the new project
28 predicted_time = model_time.predict(X_new)
29 predicted_cost = model_cost.predict(X_new)
30
31 predicted_time, predicted_cost

```

Result
`(array([238.44444444]), array([3042337.77777778]))`

Obrázek 9.3-4 LLM zvolil lineární regresi k vytvoření modelu strojového učení pro předpověď nákladů a času projektu.

Program LLM automaticky rozpoznal tabulku z přiloženého obrázku a převedl data z vizuálního formátu do tabulkového pole (obr. 9.3-4 - řádek 6). Toto pole bylo použito jako základ pro vytvoření rysů a značek, z nichž byl vytvořen model strojového učení (obr. 9.3-4 - 17.-22. řádek), který využíval lineární regresi.

Pomocí základního lineárního regresního modelu, který byl vycvičen na "extrémně malém" souboru dat, byly provedeny předpovědi pro nový hypotetický stavební projekt označený jako Projekt X. V našem

problému je tento projekt charakterizován 40 byty, 4 podlažími a úrovní složitosti 7 (obrázek 9.3-2).

Jak předpovídá lineární regresní model založený na omezeném a malém souboru údajů pro nový projekt X (obrázek 9.3-4 - řádek 24-29):

- **Délka výstavby** bude přibližně 238 dní (238.4444444).
- **Celkové výdaje** budou činit přibližně \$ 3 042 338 (3042337,777).

Pro další zkoumání hypotézy o nákladech projektu je užitečné experimentovat s různými algoritmy a metodami strojového učení. Proto předpovídejme stejné hodnoty nákladů a času pro nový projekt X na základě malého souboru historických dat pomocí algoritmu K-Nearest Neighbours (k-NN).

Předpovědi nákladů a času projektu pomocí algoritmu K-nejbližšího souseda (k-NN).

Jako další prediktor pro odhad nákladů a doby trvání nového projektu používáme algoritmus k-Nearest Neighbours (k-NN). Algoritmus k-Nearest Neighbours (k-NN) je metoda strojového učení pod dohledem (supervised machine learning) pro klasifikaci i regresi. Algoritmus k-NN jsme již dříve probírali také v souvislosti s vyhledáváním vektorových databází (obrázek 8.2-2), kde se používá k vyhledávání nejbližších vektorů (např. textů, obrázků nebo technických popisů). V tomto přístupu je každý projekt reprezentován jako bod ve vícerozměrném prostoru, kde každá dimenze odpovídá určitému atributu projektu.

V našem případě, vzhledem ke třem atributům každého projektu, je budeme reprezentovat jako body v trojrozměrném prostoru (obrázek 9.3-5). Náš připravovaný projekt X tedy bude v tomto prostoru lokalizován souřadnicemi ($x=4, y=4, z=7$). Je třeba poznamenat, že v reálných podmírkách může být počet bodů a dimenzionalita prostoru řádově větší.

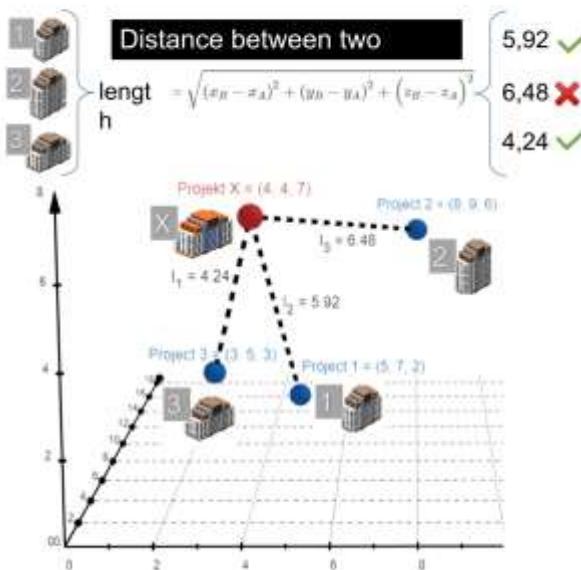
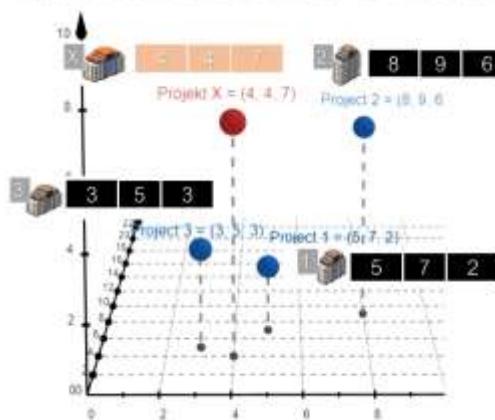
Algoritmus K-NN (k-nejbližší sousedé) funguje na základě měření vzdálenosti mezi požadovaným projektem X a projekty v tréninkové databázi. Porovnáním těchto vzdáleností algoritmus určí projekty, které jsou nejblíže bodu nového projektu X.

Pokud je například druhý projekt ($x=8, y=9, z=6$) z našeho původního souboru dat mnohem dále od X (obrázek 9.3-5) než ostatní projekty, lze jej z další analýzy vyloučit. V důsledku toho lze pro výpočty použít pouze dva ($k=2$) nejbližší projekty, na jejichž základě bude určena průměrná hodnota.

Taková metoda umožňuje prostřednictvím vyhledávání v sousedství posoudit podobnosti mezi projekty, což následně pomáhá vyvodit závěry o pravděpodobných nákladech a načasování nového projektu na základě podobných projektů, které byly realizovány dříve.

k-nearest neighbors algorithm

The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other.



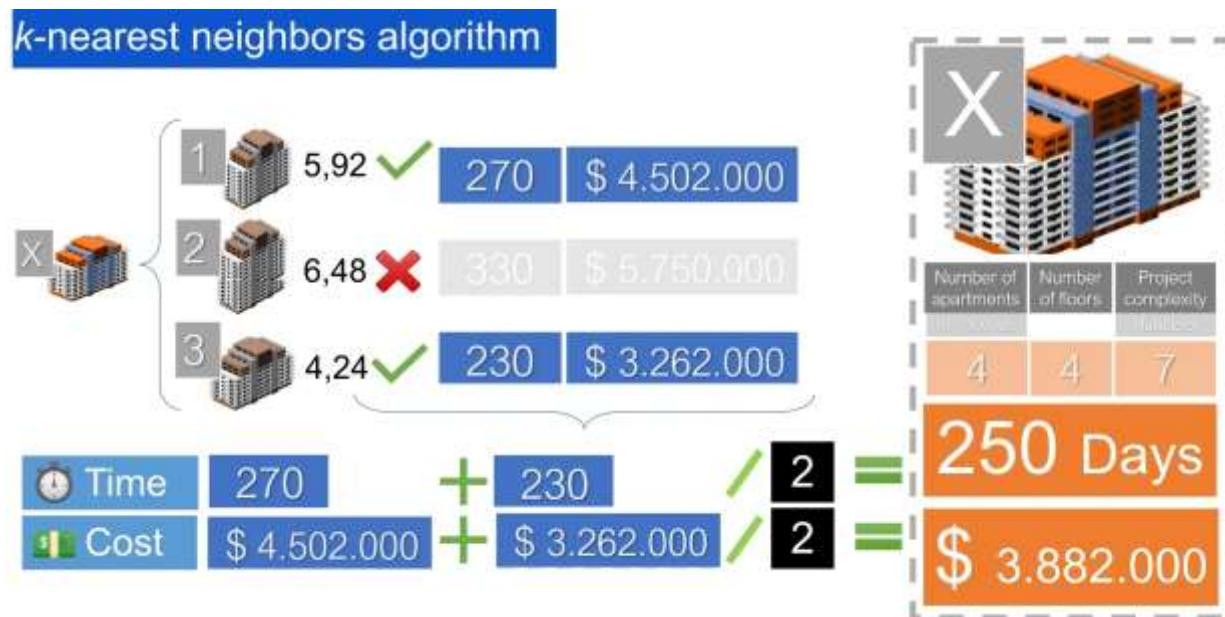
Obr. 9.3-5 V algoritmu K-NN jsou projekty reprezentovány jako body ve vícerozměrném prostoru a na základě vzdáleností jsou vybrány nejbližší projekty pro posouzení podobnosti a predikci.

Práce s k-NN zahrnuje několik klíčových kroků:

- **Příprava dat:** nejprve se načtou trénovací a testovací sady dat. Tréninková data slouží k "trénniku" algoritmu a testovací data slouží k ověření jeho účinnosti.
- **Výběr parametru K:** zvolí se číslo K, které udává, kolik nejbližších sousedů (datových bodů) se má v algoritmu zohlednit. Hodnota "K" je velmi důležitá, protože ovlivňuje výsledek.
- Proces klasifikace a regrese pro testovací data:
 - **Výpočet vzdáleností:** pro každý prvek z testovacích dat se vypočítá vzdálenost ke každému prvku z trénovacích dat (obrázek 9.3-5). K tomu lze použít různé metody měření vzdáleností, například euklidovskou vzdálenost (nejběžnější metoda), Manhattanovu vzdálenost nebo Hammingovu vzdálenost.
 - **Třídění a výběr K nejbližších sousedů:** po výpočtu vzdáleností se seřadí a vybere se K nejbližších bodů k testovanému bodu.
 - **Určení třídy nebo hodnoty testovacího bodu:** pokud se jedná o klasifikační úlohu, třída testovacího bodu se určí na základě nejčastější třídy mezi K vybranými sousedy. Pokud se jedná o regresní úlohu, vypočítá se průměr (nebo jiná míra centrální tendenze) hodnot K sousedů.
- **Dokončení procesu:** jakmile jsou všechna testovací data klasifikována nebo jsou pro ně vytvořeny předpovědi, je proces dokončen.

Algoritmus k-nejbližších sousedů (k-NN) je účinný v mnoha praktických aplikacích a je jedním z hlavních nástrojů v arzenálu odborníků na strojové učení. Tento algoritmus je oblíbený díky své jednoduchosti a efektivitě, zejména v úlohách, kde je snadné interpretovat vztahy mezi daty.

V našem příkladu byly po použití algoritmu K-nejbližší sousedé identifikovány dva projekty (z našeho malého vzorku) s nejkratší vzdáleností od projektu X (obrázek 9.3-5). Na základě těchto projektů algoritmus určí průměr jejich ceny a délky výstavby. Po analýze (obr. 9.3-6) algoritmus na základě zprůměrování nejbližších sousedů dospěl k závěru, že projekt X bude stát přibližně \$ 3 800 000 USD a jeho realizace potrvá přibližně 250 dní.



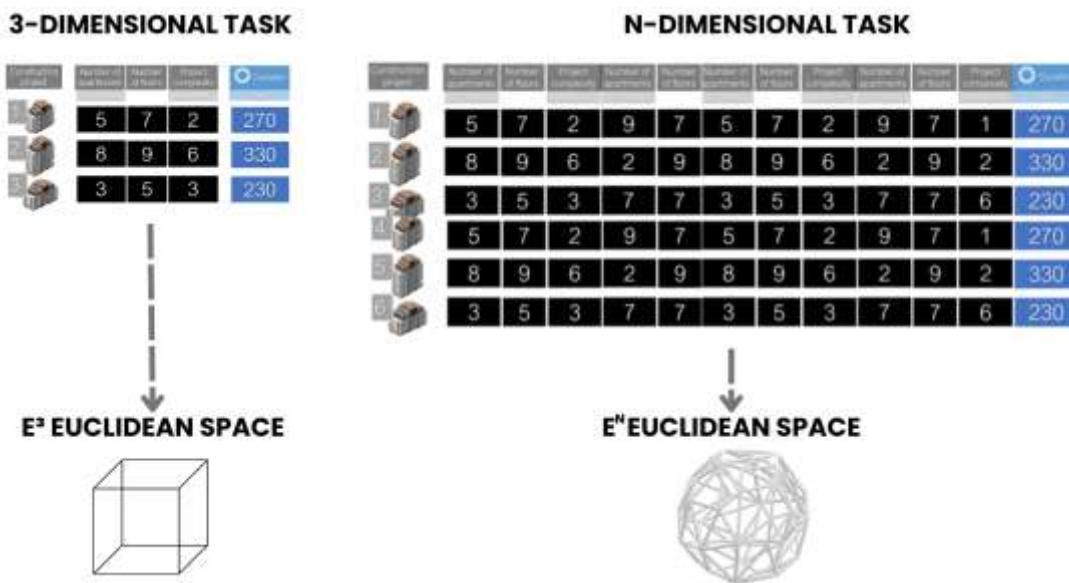
Obrázek 9.3-6 Algoritmus K-nejbližších sousedů určuje náklady a harmonogram projektu X analýzou dvou nejbližších projektů ve vzorku.

Algoritmus k-Nearest Neighbors (k-NN) je oblíbený zejména v klasifikačních a regresních úlohách, například v doporučovacích systémech, kde se používá k navrhování produktů nebo obsahu na základě preferencí podobných zájmům konkrétního uživatele. Kromě toho se k-NN hojně využívá v lékařské diagnostice ke klasifikaci typů nemocí na základě příznaků pacienta, při rozpoznávání vzorů a ve finančním sektoru k posuzování bonity zákazníků.

I s omezenými daty mohou modely strojového učení poskytovat užitečné předpovědi a výrazně zlepšit analytickou složku řízení stavebních projektů. S rozšiřováním a čištěním historických dat je možné přejít k sofistikovanějším modelům - například zohledňujícím typ stavby, lokalitu, sezónu zahájení stavby a další faktory.

V naší zjednodušené úloze jsme pro vizualizaci ve 3D prostoru použili tři atributy, ale skutečné projekty

v průměru obsahují stovky nebo tisíce atributů (viz datový soubor z kapitoly "Příklad velkých dat na bázi CAD (BIM)"), což výrazně zvyšuje dimenzionalitu prostoru a složitost reprezentace projektů jako vektorů (obrázek 9.3-7).



Obrázek 9.3-7 Ve zjednodušeném příkladu byly pro 3D -vizualizaci použity tři atributy, zatímco skutečné projekty jich mají více.

Použití různých algoritmů na stejný soubor dat pro projekt X, který má 40 bytů, 4 podlaží a úroveň složitosti 7, poskytlo různé předpovězené hodnoty. Algoritmus lineární regrese předpověděl dobu dokončení 238 dní a náklady \$ 3 042 338 (obrázek 9.3-4), zatímco algoritmus k-NN předpověděl 250 dní a \$ 3 882 000 (obrázek 9.3-6).

Přesnost předpovědí získaných pomocí modelů strojového učení, přímo závisí na objemu a kvalitě vstupních dat. Čím více projektů je zapojeno do trénování a čím úplněji a přesněji jsou reprezentovány jejich charakteristiky (atributy) a výsledky (štítky), tím vyšší je pravděpodobnost získání spolehlivých předpovědí s minimálními hodnotami chyb.

Důležitou roli v tomto procesu hrají techniky předzpracování dat, včetně:

- Normalizace, aby se prvky sjednotily do společného měřítka;
- Detekce a eliminace odlehlých hodnot, eliminace zkreslení modelu;
- Kódování kategoriálních atributů umožňující manipulaci s textovými daty;
- Doplňení chybějících hodnot, čímž se zvýší robustnost modelu.

Kromě toho se používají metody křížové validace k posouzení obecnosti modelu a jeho odolnosti vůči novým souborům dat s cílem odhalit nadměrné přizpůsobení a zvýšit spolehlivost předpovědi.

Chaos je řád, který je třeba rozluštit [160].

- José Saramago, "Dvojník"

I když se vám zdá, že chaos vašich úkolů nelze popsat formálně, měli byste vědět, že jakékoli dění ve světě a zejména stavební procesy podléhají matematickým zákonitostem, které mohou potřebovat podporu výpočtu hodnot nikoli prostřednictvím přísných vzorců, ale s pomocí statistik a historických dat.

Jak tradiční odhady oddělení kalkulace nákladů, tak modely strojového učení se nevyhnutelně potýkají s nejistotou a potenciálními zdroji chyb. Jsou-li však k dispozici dostatečně kvalitní údaje, mohou modely strojového učení vykazovat srovnatelnou a někdy i vyšší přesnost předpovědí než expertní odhady.

Strojové učení se pravděpodobně stane spolehlivým doplňkovým nástrojem pro analýzy, které mohou: zpřesnit výpočty, navrhnout alternativní scénáře a identifikovat skryté závislosti mezi parametry projektu. Takové modely si nebudou dělat nárok na univerzálnost, ale brzy zaujmou důležité místo ve výpočtech a rozhodovacích procesech. Technologie strojového učení nevyloučí účast inženýrů, odhadců a analytiků, ale naopak rozšíří jejich možnosti tím, že nabídnou další úhel pohledu založený na historických datech.

Pokud bude strojové učení vhodně integrováno do obchodních procesů stavebních společností, má potenciál stát se důležitým prvkem systému podpory rozhodování managementu - nikoli jako náhrada člověka, ale jako rozšíření jeho odborné intuice a inženýrské logiky.

Další kroky: od skladování k analýze a předpovědím

Moderní přístupy k práci s daty začínají měnit způsob rozhodování ve stavebnictví. Přechod od intuitivního hodnocení k objektivní analýze dat nejen zvyšuje přesnost, ale také otevřívá nové možnosti optimalizace procesů. Pro shrnutí této části je vhodné upozornit na hlavní praktické kroky, které vám pomohou aplikovat probírané metody v každodenních úkolech:

- Vytvoření udržitelné skladovací infrastruktury
- Pokusit se spojit různorodé dokumenty a projektová data do jednoho tabulkového modelu a agregovat klíčové informace do jednoho datového rámce pro další analýzu.
- Používejte efektivní formáty pro ukládání dat - například sloupcové formáty jako Apache Parquet namísto CSV nebo XLSX - zejména pro ty soubory, které by mohly být v budoucnu použity k trénování modelů strojového učení.
- Zavedení systému verzování dat pro sledování změn v průběhu projektu.
- Implementace nástrojů pro analýzu a automatizaci
- Začněte analyzovat historická data o projektech - podle dokumentace, modelů, odhadů - a

identifikujte vzory, trendy a anomálie.

- Hlavní procesy ETL (Extract, Transform, Load) pro automatické načítání a přípravu dat.
- Naučte se vizualizovat klíčové metriky pomocí různých bezplatných vizualizačních knihoven Pythonu.
- začít používat statistické metody a náhodný výběr vzorků k získání reprezentativních a reprodukovatelných analytických výsledků.

■ Zvyšování vyspělosti při práci s daty

- Naučte se několik základních algoritmů strojového učení na jednoduchých a přímočarých příkladech, jako je soubor dat Titanic.
- Analyzujte současné procesy a zjistěte, kde můžete přejít od rigidní logiky přičin a následků ke statistickým metodám prognózování a odhadů.
- Začněte s daty zacházet jako se strategickým přínosem, nikoli jako s vedlejším produktem: rozhodovací procesy budujte spíše na základě datových modelů než na základě konkrétních softwarových řešení.

Stavební společnosti, které si uvědomily hodnotu dat, vstupují do nové fáze vývoje, v níž konkurenční výhodu neurčuje množství zdrojů, ale rychlosť rozhodování na základě analytiky.



MAXIMÁLNÍ POHODLÍ S TIŠTĚNOU VERZÍ

Držíte v ruce bezplatnou digitální verzi knihy **Data-Driven Construction**. Pro pohodlnější práci a rychlý přístup k materiálům doporučujeme věnovat pozornost **tištěnému vydání**:



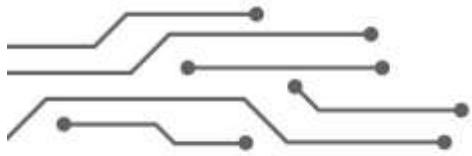
■ **Vždy po ruce:** kniha v tištěné podobě se stane spolehlivým pracovním nástrojem, který vám umožní rychle najít a použít správné vizualizace a diagramy v jakékoli pracovní situaci.

■ **Vysoko kvalitní ilustrace:** všechny obrázky a grafika v tištěném vydání jsou prezentovány v maximální kvalitě.

■ **Rychlý přístup k informacím:** snadná navigace, možnost dělat si poznámky, záložky a pracovat s knihou kdekoli.

Zakoupením plné tištěné verze knihy získáte praktický nástroj pro pohodlnou a efektivní práci s informacemi: možnost rychle využívat obrazové materiály při každodenních úkolech, rychle vyhledávat potřebná schémata a dělat si poznámky. Navíc svým nákupem podpoříte šíření otevřených znalostí.

Tištěnou verzi knihy si můžete objednat na adrese: datadrivenconstruction.io/books



X ČÁST

STAVEBNICTVÍ V DIGITÁLNÍM VĚKU. PŘÍLEŽITOSTI A VÝZVY

Poslední desátá část je komplexním pohledem na budoucnost stavebnictví v éře digitální transformace. Analyzuje posun od kauzální analýzy k práci s korelacemi velkých dat. Uvádí paralely mezi vývojem výtvarného umění a vývojem práce s daty ve stavebnictví a ukazuje, jak se odvětví posouvá od detailní kontroly k holistickému chápání procesů. Zkoumá se koncept "uberizace" stavebnictví, kdy transparentnost dat a automatizované výpočty mohou radikálně změnit tradiční obchodní modely, odstranit potřebu zprostředkovatelů a omezit možnosti spekulací. Podrobně se probírají nevyřešené otázky, jako je univerzální klasifikace prvků, což dává stavebním společnostem čas na přizpůsobení se novému prostředí. V závěru části jsou uvedena konkrétní doporučení pro utváření strategie digitální transformace, která zahrnuje analýzu slabých míst a rozšiřování služeb, aby bylo možné udržet konkurenceschopnost v měnícím se odvětví.

KAPITOLA 10.1.

STRATEGIE PŘEŽITÍ: BUDOVÁNÍ KONKURENČNÍCH VÝHOD

Korelace místo výpočtů: budoucnost stavební analytiky

Díky rychlé digitalizaci informací (obrázek 1.1-5) prochází moderní stavebnictví zásadní proměnou, kdy se data stávají nejen nástrojem, ale strategickým aktivem, které může zásadně změnit tradiční přístupy k řízení projektů a podniků.

Po tisíce let se stavební činnosti opíraly o deterministické metody - přesné výpočty, detailování a přesnou kontrolu parametrů. V prvních stoletích našeho letopočtu používali římskí inženýři matematické principy při stavbě akvaduktů a mostů. Ve středověku architekti usilovali o ideální proporce gotických katedrál a v industrializaci 20. století se vytvořily systémy standardizovaných norem a předpisů, které se staly základem pro hromadnou výstavbu.

Vektor vývoje se dnes posouvá od hledání striktně příčinných vztahů k pravděpodobnostní analýze, hledání korelací a skrytých vzorců. Odvětví vstupuje do nové fáze - klíčovým zdrojem se stávají data a na nich založená analytika nahrazuje intuitivní a lokálně optimalizované přístupy.



Obr. 10.1-1 Skrytý potenciál konstrukčních dat: existující kalkulace ve firmě jsou jen špičkou ledovce, kterou má management k dispozici pro analýzu.

Podnikový informační systém je jako ledovec (obr. 10.1-1): vedení podniku vidí jen malou část potenciálu dat, zatímco hlavní hodnota je ukryta v hlubinách. Je důležité hodnotit data nejen z hlediska jejich současného využití, ale také z hlediska příležitostí, které se díky nim v budoucnu otevřou. Právě ty společnosti, které se naučí z dat získávat skryté vzorce a vytvářet nové poznatky, si vytvoří trvalou konkurenční výhodu

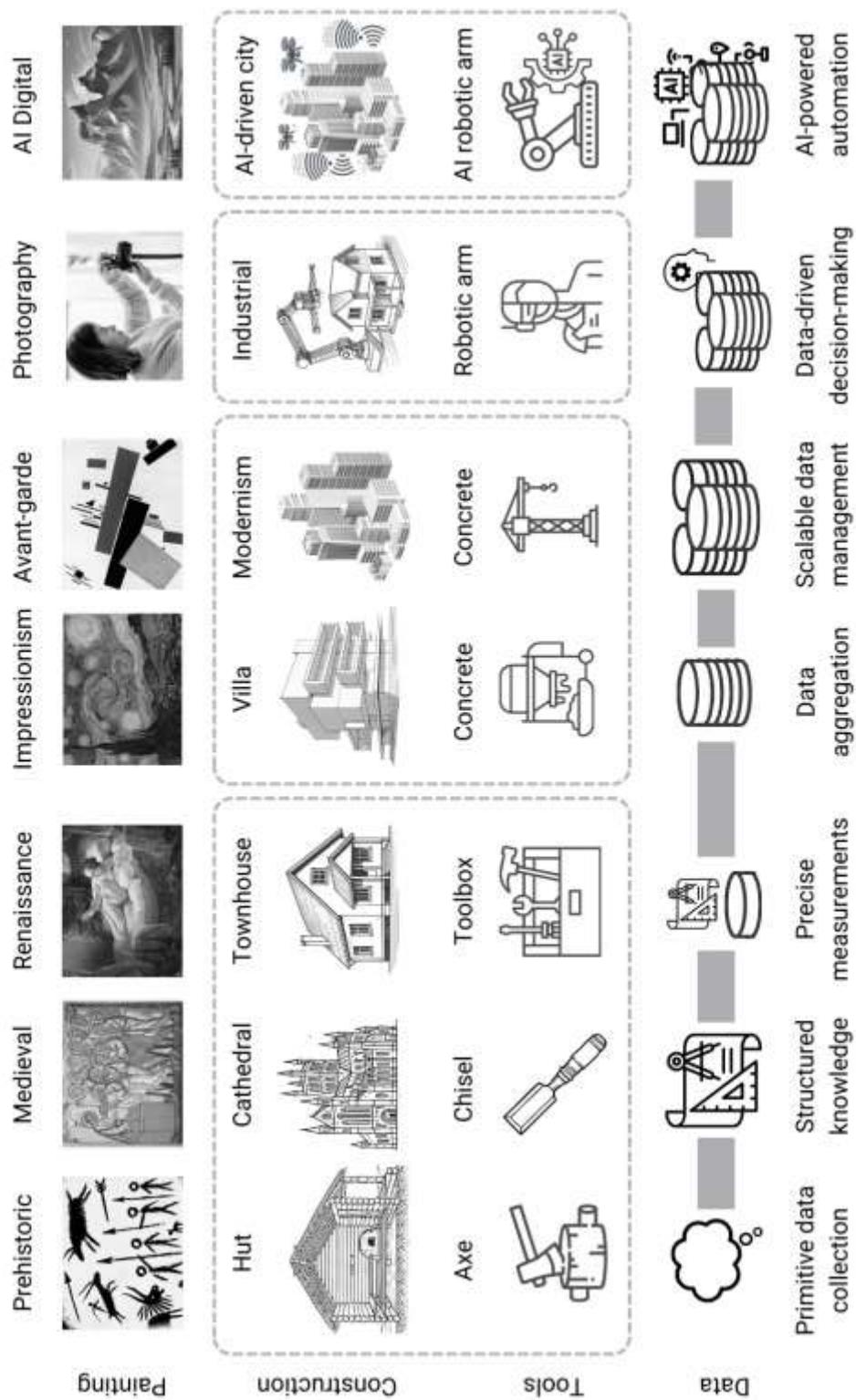
Hledání skrytých vzorců a hledání smyslu dat není jen o práci s čísly, ale o tvůrčím procesu, který vyžaduje abstraktní myšlení a schopnost vidět za různými prvky celkový obraz. V tomto smyslu lze vývoj práce s daty přirovnat k vývoji výtvarného umění (obrázek 10.1-2).

Vývoj stavebnictví je pozoruhodně podobný vývoji výtvarného umění. V obou případech lidstvo pokročilo od primitivních metod k sofistikovaným technologiím vizualizace a analýzy. V pravěku lidé používali k řešení každodenních úkolů jeskynní kresby a primitivní nástroje. Během středověku a renesance se úroveň sofistikovanosti architektury a umění dramaticky zvýšila. Do raného středověku se stavební nástroje vyvinuly z jednoduché sekery k rozsáhlým sadám nástrojů symbolizujícím růst technických znalostí.

Doba realismu znamenala první revoluci ve výtvarném umění: umělci se naučili reprodukovat i ty nejmenší detaily a dosáhnout maximální věrohodnosti. Ve stavebnictví byly protějškem tohoto období přesné inženýrské techniky, podrobné výkresy a přísně regulované výpočty, které se staly základem projekční praxe na celá staletí.

Později impresionismus změnil samotné vnímání umělecké reality: namísto doslovného ztvárnění tvaru začali umělci zachycovat náladu, světlo a dynamiku a snažili se spíše o vyjádření celkového dojmu než o absolutní přesnost. Podobně i strojové učení v oblasti analýzy budov se posouvá od rigidních logických modelů k rozpoznávání vzorů a pravděpodobnostních vzorců, které umožňují "vidět" skryté závislosti v datech, jež jsou při klasické analýze nedostupné. Tento přístup je ozvěnou myšlenek minimalismu a funkčnosti Bauhausu, kde je význam (funkce) důležitější než forma. Bauhaus se snažil odstranit nadbytečné, opustit ornamenty ve prospěch přehlednosti, utilitarismu a masové přitažlivosti. Věci musely být srozumitelné a užitečné, bez zbytečností - estetika se zrodila z logiky designu a účelu.

S příchodem fotografie na konci 19. století získalo umění nový nástroj pro nebývale přesné zachycení reality a převrátilo postoj k výtvarnému umění naruby. Podobně i ve stavebnictví vede průmyslová revoluce v 21. století k využívání robotických technologií, laserů, internetu věcí, RFID a konceptů, jako je Connected Construction, kde se sběr jednotlivých parametrů vyvinul ve škálovatelné inteligentní zachycení celé reality staveniště.



Obr. 10.1-2 Éra výtvarného vývoje odpovídá vývoji přístupů k práci s daty ve stavebnictví.

Stejně jako výtvarné umění dnes zažívá přehodnocení s příchodem nástrojů umělé inteligence a LLM,

zažívá stavebnictví další skok: inteligentní systémy řízené umělou inteligencí (AI), LLM chaty umožňují předvídat, optimalizovat a generovat řešení s minimálním zásahem člověka.

Úloha dat při navrhování a řízení se radikálně změnila. Zatímco dříve se znalosti předávaly ústně a měly empirický charakter - stejně jako se realita až do 19. století zachycovala ručně malovanými obrazy -, dnes se klade důraz na kompletní digitální zachycení "obrazu" stavby. Pomocí algoritmů strojového učení se tento digitální obraz mění na impresionistickou reprezentaci stavební reality - nikoli na přesnou repliku, ale na zobecněné, pravděpodobnostní chápání procesů.

Rychle se blížíme do éry, kdy budou procesy navrhování, výstavby a provozu budov nejen rozšířeny, ale z velké části řízeny systémy umělé inteligence. Stejně jako se moderní digitální umění vytváří bez štětce - pomocí textových podnětů a generativních modelů - budou architektonická a inženýrská řešení budoucnosti utvářena na základě klíčových dotazů a parametrů zadaných uživatelem.

V 21. století se přístup k datům, jejich interpretace a kvalita analytických dat stávají nezbytnými pro úspěch projektu. A hodnota dat není dána jejich objemem, ale schopností odborníků je analyzovat, ověřit a proměnit v akci.

Přístup založený na datech ve stavebnictví: nová úroveň infrastruktury

Každý takový technologický skok v historii lidstva přinesl zásadní změny v ekonomice a společnosti. Dnes jsme svědky nové vlny transformace, která je svým rozsahem srovnatelná s průmyslovou revolucí 19. století. Zatímco však před sto lety byly hlavním motorem změn mechanické síly a energetické technologie, nyní jsou to data a umělá inteligence.

Agenti strojového učení, LLM a AI mění samotnou podstatu aplikací a činí tradiční softwarové zásobníky (o nichž pojednává druhá část knihy) zbytečnými (obr. 2.2-3). Veškerá datová logika je soustředěna v agentech AI, nikoli v pevně zakódovaných obchodních pravidlech (obr. 2.2-4).

V době dat se zásadně mění tradiční pohled na aplikace. Přecházíme k modelu, kdy objemné, modulární podnikové systémy nevyhnutelně ustoupí otevřeným, lehkým a přizpůsobeným řešením.

V budoucnu zůstane pouze základní datová struktura a veškerá interakce s ní bude probíhat prostřednictvím agentů pracujících přímo s databází. Skutečně věřím, že celý aplikační stack zmizí, protože ho umělá inteligence při přímé interakci s podkladovou databází jednoduše nebude potřebovat. Celou svou kariéru jsem strávil prací v SaaS - budoval jsem firmy, pracoval jsem v nich, a abych byl upřímný, pravděpodobně bych teď novou firmu SaaS nespustil. A pravděpodobně bych teď do společnosti SaaS neinvestoval. Situace je příliš nejistá. To neznamená, že v budoucnu nebudou existovat softwarové společnosti, jen budou vypadat úplně jinak. Budoucí systémy budou databáze s obchodní logikou přenesenou do agentů [AI]. Tito agenti budou pracovat s více datovými úložišti najednou, nebudou omezeni na jedinou databázi. Veškerá logika se přesune do vrstvy AI [46].

- Matthew Berman, generální ředitel společnosti Forward Future

Klíčovým rozdílem nového paradigmatu je minimalizace technologického balastu. Místo monumentálních složitých a uzavřených softwarových systémů získáme flexibilní, otevřené a rychle přizpůsobitelné moduly, které doslova "žijí" uvnitř datového toku (obr. 7.4-1 - Apache Airflow, NiFi). Architektura budoucího řízení procesů předpokládá využití mikroaplikací - kompaktních, účelových nástrojů, které se zásadně liší od masivních a uzavřených systémů ERP, PMIS, CDE, CAFM. Noví agenti budou co nejvíce adaptivní, integrovaní a přizpůsobení konkrétnímu podniku (např. Low-Code/No-Code Obr. 7.4-6).

Veškerá obchodní logika bude přecházet na tyto agenty [AI] a tito agenti budou provádět operace CRUD [Create, Read, Update, and Delete] na více úložištích, což znamená, že nebudou rozlišovat, který backend se používá. Budou aktualizovat více databází a veškerá logika skončí v takzvané AI úrovni. A jakmile se veškerá logika ocítne na úrovni AI, lidé začnou backendy vyměňovat. Již nyní vidíme poměrně vysoké procento výtěžství na trhu v oblasti backendů Dynamics a používání agentů a budeme se tímto směrem agresivně ubírat a snažit se vše spojit dohromady. Ať už v oblasti zákaznických služeb, nebo v jiných oblastech, například nejen CRM, ale i naše finanční a provozní řešení. Lidé totiž chtějí více obchodních aplikací řízených umělou inteligencí, kde logickou vrstvu může řídit umělá inteligence a AI agenti. [...]. Jednou z nejzajímavějších věcí je pro mě Excel s Pythonem, což je srovnatelné s GitHubem s Copilotem. Tedy to, co jsme udělali: teď, když máte Excel, měli byste ho prostě otevřít, spustit Copilot a začít si s ním hrát. Už to není jen porozumění dostupným číslům - sám vytvoří plán. Stejně jako pracovní plocha GitHub Copilot vytvoří plán a pak ho provede, je to jako datový analytik, který používá Excel jako nástroj pro vizualizaci řádků a sloupců pro analýzu. Copilot tedy používá Excel jako nástroj se všemi jeho možnostmi, protože umí generovat data a má interpret jazyka Python.

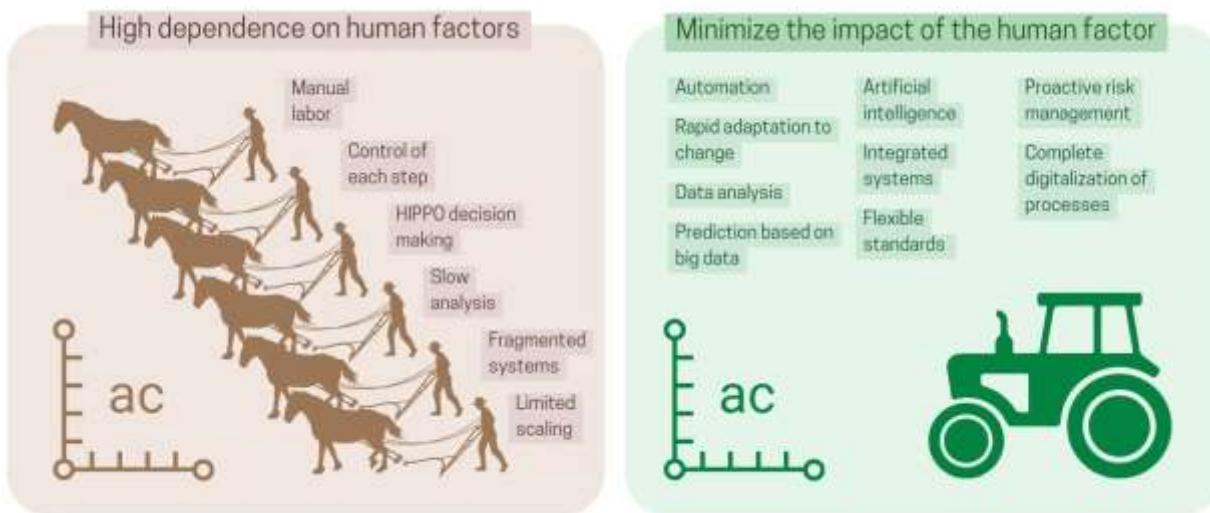
- Satya Nadella, generální ředitel společnosti Microsoft, rozhovor s kanálem BG2 prosinec 2024. [28]

Proměna, které jsme svědky v logice kancelářských aplikací - přechod od modulárních uzavřených systémů k agentům umělé inteligence pracujícím přímo s otevřenými daty - je pouze součástí mnohem

rozsáhlejšího procesu. Nejde jen o změnu rozhraní nebo softwarové architektury: změny ovlivní základní principy organizace práce, rozhodování a řízení podniku. Ve stavebnictví to povede k logice založené na datech, v níž se data stanou ústředním prvkem procesů od projektování až po řízení zdrojů a monitorování výstavby.

Digitální kancelář nové generace: jak AI mění pracovní prostor

Podobnou technologickou revoluci zažívalo lidstvo již téměř před sto lety. Přechod od parních strojů k elektromotorům trval více než čtyři desetiletí, ale nakonec urychlil nebývalý růst produktivity - především díky decentralizaci energetických kapacit a flexibilitě nových řešení. Tento posun nejen změnil běh dějin a přesunul většinu obyvatelstva z venkova do měst, ale také položil základy moderní ekonomiky. Historie technologií je cestou od fyzické práce k automatizaci a inteligentním systémům. Stejně jako traktor nahradil desítky kypřičů půdy, moderní digitální technologie nahrazují tradiční kancelářské metody řízení staveb (obrázek 10.1-3). Ještě na počátku 20. století většina světové populace obdělávala půdu ručně, dokud ve 30. letech 20. století nezačala mechanizace práce pomocí strojů a traktorů.



Obr. 10.1-3 Stejně jako traktor nahradil na počátku 20. století desítky lidí, nahradí strojové učení v 21. století tradiční metody řízení podniků a projektů.

Stejně jako lidstvo před sto lety přešlo od obdělávání jednotlivých pozemků pomocí primitivních nástrojů k velkoplošnému zemědělství s využitím strojů, přecházíme dnes od zpracování různorodých "sil" informací k práci s datovými poli pomocí výkonných "traktorů" - ETL-pipeline a algoritmů umělé inteligence.

Jsme na prahu podobného skoku - ale v digitální rovině: od tradičního manuálního řízení podniku k modelům založeným na datech.

Cesta k plnlohodnotné architektuře založené na datech bude vyžadovat čas, investice a organizační úsilí. Tato cesta však otevírá cestu nejen k postupnému zlepšování, ale ke kvalitativnímu skoku směrem k větší efektivitě, transparentnosti a řiditelnosti stavebních procesů. To vše je podmíněno

systematickým zaváděním digitálních nástrojů a opuštěním zastaralých obchodních praktik.

Parametrizace úloh, ETL, LLM, komponenty IoT, RFID, tokenizace, big **data** a strojové učení změní tradiční stavebnictví na **datově řízené stavebnictví**, kde bude každý detail projektu a stavební činnosti řízen a optimalizován pomocí dat.

Analýza informací dříve zabrala tisíce člověkohodin. Nyní tyto úkoly provádějí algoritmy a LLM, které pomocí výzev mění různorodé soubory dat na strategické zdroje. Ve světě technologií se děje totéž, co se stalo v zemědělství: přecházíme od motyk k automatizovanému agrotechnickému podnikání. Stejně tak kancelářská práce ve stavebnictví - od souborů v Excelu a ručního summarizování - přechází k inteligentnímu systému, kde se data shromažďují, čistí, strukturují a mění v poznatky.

Již dnes by firmy měly začít "pěstovat" informační pole pomocí kvalitního sběru dat a strukturování informací, "hnojit" je čisticími a normalizačními nástroji a následně je "sklízet" formou prediktivní analýzy a automatizovaných řešení. Jestliže moderní zemědělec se strojem dokáže nahradit sto kypřičů půdy, inteligentní algoritmy budou schopny sejmout rutinu ze zaměstnanců a převést je do role strategických správců informačních toků.

Je však důležité si uvědomit, že vytvoření skutečně datově orientované -organizace není rychlý proces. Jde o dlouhodobé strategické směřování, podobné vytváření nového místa pro pěstování nového lesa (obr. 1.2-5) systémů, kde každý "strom" v tomto ekosystému je samostatným procesem, kompetencí nebo nástrojem, jehož růst a rozvoj vyžaduje čas. Stejně jako v případě skutečného lesa závisí úspěch nejen na kvalitě sadebního materiálu (technologie), ale také na půdě (firemní kultura), klimatu (podnikatelské prostředí) a péči (systémový přístup).

Společnosti se již nebudou moci spoléhat pouze na uzavřená, out-of-the-box řešení. Na rozdíl od předchozích etap technologického vývoje je nepravděpodobné, že by současný přechod - k otevřenému přístupu k datům, umělé inteligenci a otevřenému zdrojovému kódu - podporovali velcí dodavatelé, protože přímo ohrožuje jejich zavedené obchodní modely a hlavní zdroje příjmů.

Jak ukázala studie Harvard Business School [40], o které již bylo pojednáno v kapitole o čtvrté a páté technologické revoluci, náklady na vytvoření nejpoužívanějších Open Source řešení od nuly pro všechny společnosti by činily přibližně 4,15 miliardy dolarů. Pokud si však představíme, že by si každá společnost využila vlastní alternativy bez přístupu k existujícím nástrojům Open Source, což se děje v posledních desetiletích, mohly by celkové náklady podniků dosáhnout kolosálních 8,8 bilionu dolarů - to je cena iracionální poptávky, na kterou lze ocenit trh se softwarem.

Technologický pokrok nevyhnutelně povede k přehodnocení zavedených obchodních modelů. Zatímco dříve mohly společnosti vydělávat na složitých, neprůhledných procesech a uzavřených datech, s rozvojem umělé inteligence a analytiky je tento přístup stále méně schůdný.

V důsledku demokratizace přístupu k datům a nástrojům se může tradiční trh s prodejem softwaru výrazně zmenšit. Současně však poroste nový trh - trh digitálních odborných znalostí, přizpůsobení, integrace a návrhu řešení. Zde nebude hodnota plynout z prodeje licencí, ale ze schopnosti vytvářet flexibilní, otevřené a přizpůsobivé digitální procesy. Stejně jako elektrifikace a nástup traktorů daly vzniknout novým průmyslovým odvětvím, tak i aplikace velkých dat, umělé inteligence a LLM otevře zcela nové obzory pro podnikání ve stavebnictví, což bude vyžadovat nejen technologické investice, ale také hlubokou proměnu myšlení, procesů a organizačních struktur. A ty společnosti a odborníci, kteří si to uvědomí a začnou jednat již dnes, budou lídry zítřka.

Ve světě, kde se otevřená data stávají hlavním přínosem, bude dostupnost informací měnit pravidla hry. Investoři, klienti a regulační orgány budou stále více vyžadovat transparentnost a algoritmy strojového učení budou schopny automaticky identifikovat nesrovnanosti v odhadech, časových plánech a nákladech. To vytváří podmínky pro novou etapu digitální transformace, která nás postupně vede k "uberizaci" stavebnictví.

Otevřená data a uberizace jsou hrozbou pro stávající stavební firmy

Stavebnictví se stává procesem řízení informací. Čím přesnější, kvalitnější a úplnější jsou data, tím efektivnější je navrhování, výpočty, odhady nákladů, výstavba a provoz budov. V budoucnu nebude klíčovým zdrojem jeřáb, beton a výztuž, ale schopnost shromažďovat, analyzovat a využívat informace.

Klienti stavebních společností - investoři a klienti financující výstavbu - budou v budoucnu nevyhnutelně využívat hodnoty otevřených dat a analytiky historických dat. Otevřou se tak možnosti automatizace výpočtu harmonogramu projektu a nákladů, aniž by se stavební společnosti musely zabývat otázkami kalkulace nákladů, což pomůže kontrolovat náklady a rychleji identifikovat nadbytečné náklady.

Představte si staveniště, kde laserové skenery, kvadrokoptéry a fotogrammetrické systémy shromažďují přesná data o množství použitého betonu v reálném čase. Tyto informace jsou automaticky převedeny do jednoduchých plochých MESH -modelů s metadaty, čímž se obejdou těžkopádné systémy CAD (BIM), bez závislosti na složitých geometrických jádrex, ERP nebo PMIS. Tato data shromážděná ze stavby jsou centrálně přenášena do jediného strukturovaného úložiště, které je k dispozici klientovi pro nezávislou analýzu, kam se nahrávají reálné ceny z různých stavebních obchodů s parametry od míry úvěrového financování až po dynamicky se měnící faktory, jako jsou povětrnostní podmínky, burzovní kotace stavebních materiálů, logistické tarify a statistické sezónní výkyvy cen práce. Za těchto podmínek jsou okamžitě zřejmě jakékoli rozdíly mezi projektem a skutečnými objemy materiálu, což znemožňuje manipulaci s odhadů jak ve fázi projektování, tak při předání projektu. Výsledkem je, že transparentnost stavebního procesu není zajištěna armádou dozorců a manažerů, ale objektivními digitálními daty, která minimalizují lidský faktor a možnost spekulací.

V budoucnu budou tento druh kontroly dat provádět spíše správci dat na straně zákazníka (obr. 1.2-4 Správce CQMS). To platí zejména pro kalkulace a odhady projektů: tam, kde dříve existovalo celé oddělení odhadců, budou již zítra existovat nástroje strojového učení a prognostické nástroje, které budou stavebním firmám stanovovat cenové limity, do nichž se vejdu.

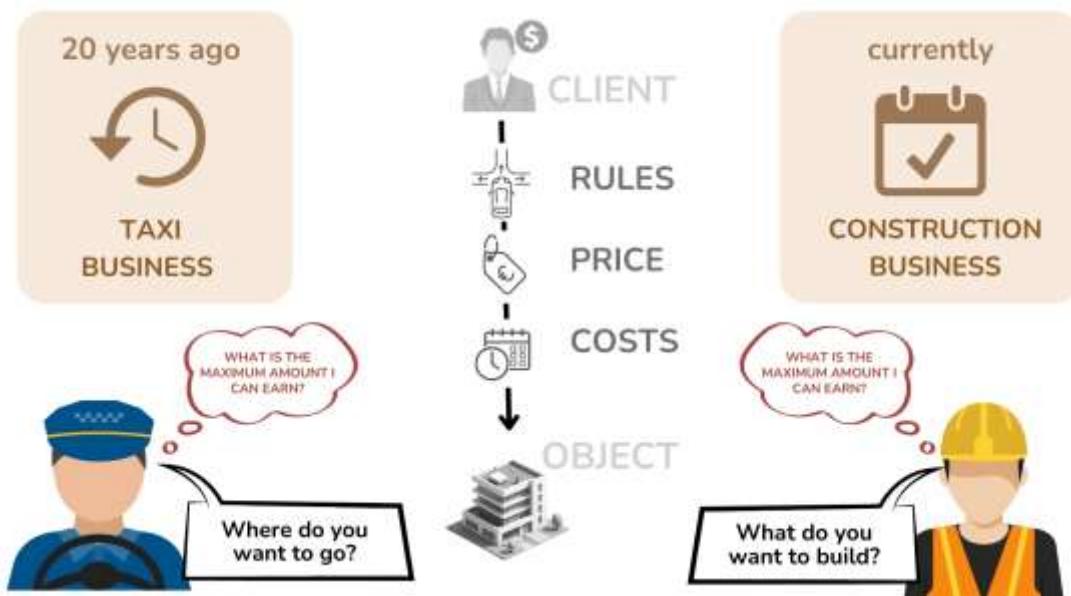
*Vzhledem k roztríštěnosti [stavebního] průmyslu, kde většinu systémů a subsystémů dodávají malé a střední podniky, musí digitální strategie vycházet od zákazníka.
Zákazníci musí vytvořit podmínky a mechanismy pro uvolnění digitálních schopností dodavatelského řetězce [20].*

- Andrew Davis a Giuliano Denicol, Accenture "Vytváření větší hodnoty prostřednictvím kapitálových projektů"

Taková otevřenosť a transparentnosť dat představuje hrozbu pro stavební společnosti, které jsou zvyklé vydělávat na neprůhledných procesech a matoucích výkazech, kde se za složitými a uzavřenými formáty a modulárními proprietárními datovými platformami mohou skrývat spekulace a skryté náklady. Proto stavební společnosti, stejně jako dodavatelé propagující open source řešení, pravděpodobně nebudou mít zájem plně implementovat otevřená data do svých obchodních procesů. Pokud jsou data dostupná a snadno zpracovatelná pro zákazníka, lze je automaticky kontrolovat, což vyloučí možnost nadhodnocování objemů a manipulace s odhady.

Podle zprávy Světového ekonomického fóra "Formování budoucnosti stavebnictví" (2016) [5] zůstává jednou z klíčových výzev pro toto odvětví pasivní role klienta. Nicméně jsou to právě zákazníci, kteří by měli převzít větší odpovědnost za výsledek projektů - od včasného plánování, přes výběr udržitelných modelů interakce až po sledování výkonnosti. Bez aktivní účasti vlastníků projektů není systémová transformace stavebnictví možná.

Ztráta kontroly nad objemem a kalkulací nákladů již za posledních 20 let změnila jiná odvětví a umožnila zákazníkům přímo, bez zprostředkovatelů, brzdit jejich cíle. Digitalizace a transparentnosť dat proměnily mnoho tradičních obchodních modelů, například taxikáře s příchodem Uberu (obrázek 10.1-4), hoteliéry s příchodem Airbnb a maloobchodníky a obchody s nástupem Amazonu a banky s nástupem nebankovních a decentralizovaných fintech ekosystémů, kde přímý přístup k informacím a automatizace kalkulací času a nákladů výrazně omezily roli zprostředkovatelů.



Obr. 10.1-4 Stavebnictví bude čelit überizaci, které museli před deseti lety čelit taxikáři, hoteliéři a maloobchodníci.

Proces demokratizace přístupu k datům a nástrojům pro jejich zpracování je nevyhnutelný a postupem času se otevřená data o všech součástech projektu stanou požadavkem zákazníka a novým standardem. Proto budou otázky zavádění otevřených formátů a transparentních výpočtů prosazovat investoři, zákazníci, banky a fondy soukromého kapitálu (private equity) - tedy ti, kteří jsou konečnými uživateli vybudovaných objektů a objekt pak provozuje desítky let.

Velcí investoři, klienti a banky již požadují transparentnost ve stavebnictví. Podle studie společnosti Accenture "Vytváření větší hodnoty prostřednictvím investičních projektů" (2020) [20] se transparentní a spolehlivé údaje stávají rozhodujícím faktorem pro investiční rozhodování ve stavebnictví. Jak upozorňují odborníci, důvěryhodné a efektivní řízení projektů není možné bez transparentnosti, zejména v době krizí. Kromě toho se vlastníci majetku a dodavatelé stále častěji přiklánějí ke smlouvám, které motivují ke sdílení dat a ke spolupráci v oblasti analytiky, což odráží rostoucí požadavky investorů, bank a regulačních orgánů na odpovědnost a transparentnost.

Pohyb investora, zákazníka od nápadu k hotové stavbě, se v budoucnu bude podobat cestování na autopilotu - bez řidiče v podobě stavební firmy, slibuje se stát nezávislým na spekulacích a nejistotě.

Éra otevřených dat a automatizace nevyhnutelně změní stavebnictví, stejně jako už změnila bankovnictví, obchod, zemědělství a logistiku. V těchto odvětvích ustupuje úloha zprostředkovatelů a tradiční způsoby podnikání automatizaci a robotizaci a nezbývá prostor pro neoprávněné přirážky a spekulace.

Data a procesy ve všech ekonomických činnostech člověka se nijak neliší od toho, s čím se potýkají

odborníci ve stavebnictví. Z dlouhodobého hlediska mohou stavební firmy, které dnes dominují trhu tím, že určují ceny a standardy kvality služeb, ztratit svou roli klíčového prostředníka mezi zákazníkem a jeho stavebním projektem.

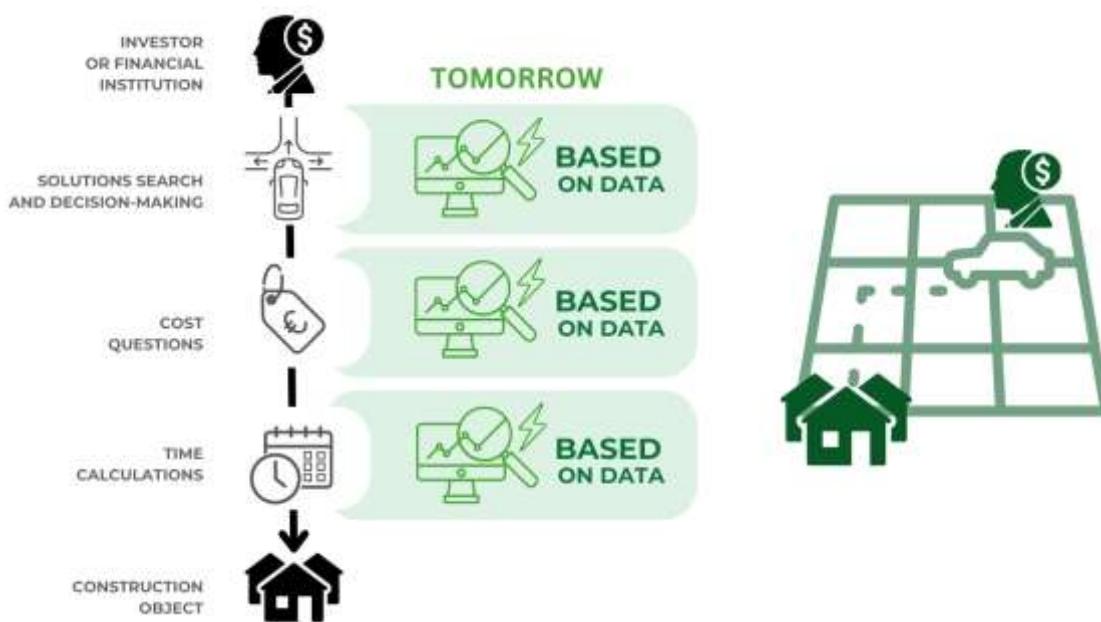
Nevyřešené problémy uberizace jako poslední šance využít čas k transformaci

Vraťme se však k realitě stavebnictví. Zatímco v některých odvětvích ekonomiky se objevují samořízená auta, decentralizované finanční systémy a řešení založená na umělé inteligenci, značná část stavebních firem stále zůstává papírovými organizacemi, kde se klíčová rozhodnutí přijímají spíše na základě intuice a zkušeností jednotlivých specialistů.

V tomto paradigmatu lze moderní stavební firmu přirovnat k 20 let staré taxislužbě, která řídí zdroje, trasy a dodací lhůty a je zodpovědná za načasování a náklady "cesty" - od nápadu na projekt (logistický a montážní proces) až po dodání projektu. Stejně jako kdysi GPS (ve stavebnictví IoT, RFID) a algoritmy strojového učení v kalkulacích času/nákladů proměnily oblast dopravy, mají data, algoritmy a AI-agenti potenciál proměnit řízení stavebnictví - od intuitivního hodnocení k prediktivním, řízeným modelům. Za posledních 20 let se v mnoha odvětvích - ve finančnictví, zemědělství, maloobchodě a logistice - postupně vytratila možnost spekulovat díky nepřehlednosti dat. Ceny, náklady na dodávky nebo finanční transakce se vypočítávají automaticky a statisticky podloženě - během několika sekund na digitálních platformách.

Při pohledu do budoucna si stavební společnosti musí uvědomit, že demokratizace přístupu k datům a nástrojům pro jejich analýzu naruší tradiční přístup k odhadování nákladů a časového harmonogramu projektů a odstraní možnost spekulovat na základě neprůhledných údajů o objemu a ceně.

Stejně jako jízda po regulované silnici bez zásahu řidiče budou stavební procesy budoucnosti stále více připomínat "uberizovaný" systém - s automatizovaným odhadem času a nákladů, transparentním směrováním úkolů a minimální závislostí na lidském faktoru. Tím se změní samotná povaha "cesty" od nápadu k realizaci - bude předvídatelnější, lépe zvládnutelná a řízená daty.



Obrázek 10.1-5 Náklady a "cestovní" čas během výstavby budou stanoveny pomocí strojového učení a statistických nástrojů.

S postupným zaváděním nových předpisů a požadavků téměř ve všech zemích světa, které ukládají povinnost předávat klientům nebo bankám financujícím stavební projekty modely CAD- (BIM), mají klient a zákazník možnost nezávisle zajistit transparentnost kalkulací nákladů a rozsahu. To je důležité zejména pro velké odběratele a investory, kteří mají dostatečné kompetence a nástroje k rychlé analýze objemů a sledování tržních cen. Pro společnosti realizující velké standardní projekty - obchody, kancelářské budovy, obytné komplexy - se tyto postupy stávají standardem.

S tím, jak se informační obsah modelů stává úplnějším a standardizovanějším, prakticky mizí možnost manipulace a spekulací. Digitální transformace postupně mění pravidla hry ve stavebnictví a společnosti, které se těmto změnám nepřizpůsobí, mohou čelit vážným problémům.

Rostoucí konkurence, narušení technologického rozvoje a snižující se marže mohou mít dopad na udržitelnost podnikání. Vzhledem k omezené likviditě se stále více hráčů v oboru obrací k automatizaci, analytice a datovým technologiím jako ke způsobu, jak zvýšit efektivitu a transparentnost procesů. Tyto nástroje se stávají důležitým zdrojem pro udržení konkurenční schopnosti v měnícím se ekonomickém prostředí.

Možná bychom neměli čekat, až nás vnější okolnosti donutí k naléhavým opatřením - mnohem efektivnější je začít se připravovat již dnes posílením digitálních kompetencí, zaváděním moderních řešení a budováním kultury zaměřené na data.

Jednou z posledních klíčových technologických překážek rozsáhlé digitální transformace stavebnictví, která se v příštích letech dotkne každé společnosti, je problém automatické klasifikace prvků stavebních projektů.

Bez spolehlivé, přesné a škálovatelné klasifikace není možné vytvořit základ pro plnohodnotnou analýzu, automatizaci procesů a řízení životního cyklu pomocí AI a prediktivních modelů. Dokud bude klasifikace objektů stále záviset na manuální interpretaci zkušenými odborníky - mistry, projektanty, odhadci - má stavebnictví stále ještě příležitost. Tento čas lze využít k přípravě na nevyhnutelné změny: rostoucí požadavky na transparentnost, demokratizaci nástrojů a dat a nástup automatických klasifikačních systémů, které radikálně změní pravidla hry.

Úloha automatické klasifikace prvků stavebního světa je svou složitostí srovnatelná s rozpoznáváním objektů v bezpilotních systémech řízení, což je jedna z hlavních výzev. Představme si bezpilotní automobil jedoucí z bodu A do bodu B (obr. 10.1-5). Současné systémy automatického řízení se potýkají s problémem klasifikace objektů, které jsou rozpoznávány lidary a kamerami. Nestačí, aby automobil pouze "viděl" překážku nebo orientační bod, ale musí být schopen rozpoznat, zda se jedná o chodce, dopravní značku nebo odpadkový koš.

Podobnému zásadnímu problému čelí celé stavebnictví. Prvky projektu - jako jsou okna, dveře nebo sloupy - mohou být zaznamenány v dokumentaci, zobrazeny v modelech CAD, vyfotografovány na staveništi nebo rozpoznány v mračnech bodů z laserového skenování. Jejich vizuální nebo hrubé geometrické rozpoznání však nestačí k vytvoření skutečně automatizovaného systému řízení projektů. Je nutné zajistit, aby každý prvek byl přesně a důsledně zařazen do typu, který je jednoznačně identifikovatelný ve všech následných procesech - od odhadů a specifikací až po logistiku, řízení zásob a především provoz (obr. 4.2-6).

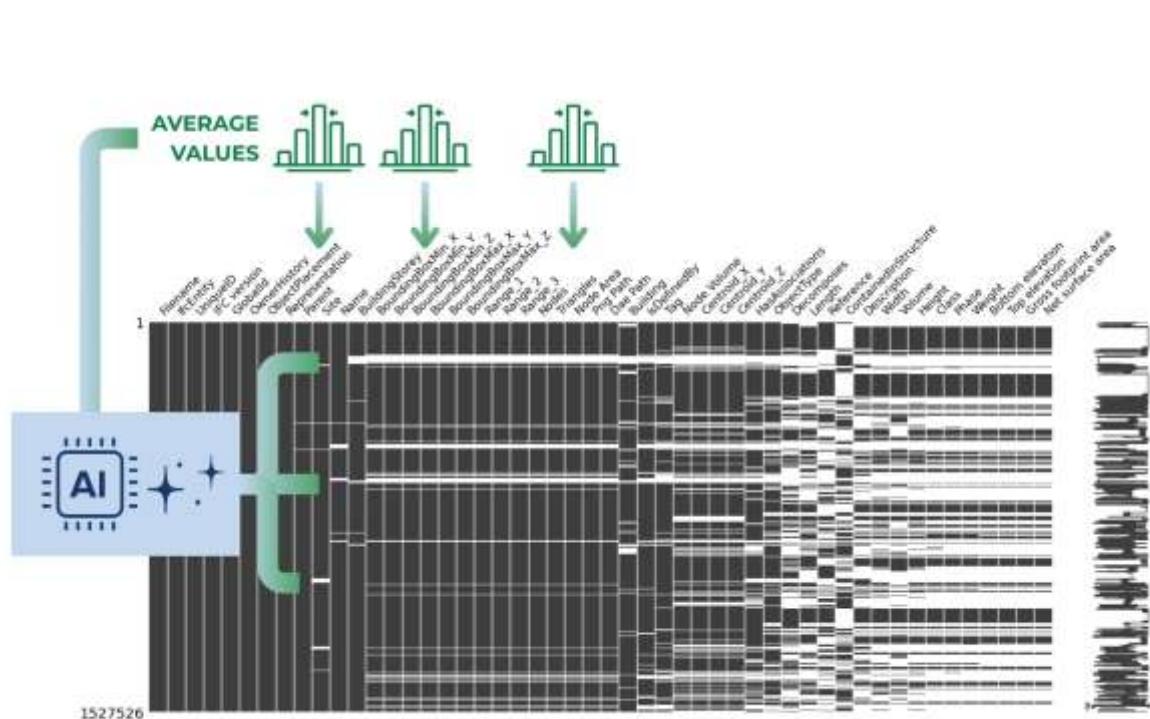
Právě v této fázi - při přechodu od rozpoznání ke smysluplné klasifikaci - vzniká jedna z klíčových překážek. I když jsou digitální systémy technicky schopné rozlišovat a identifikovat objekty v modelech a na staveništi, hlavní potíž spočívá ve správné a kontextově stabilní definici typu prvku pro různá softwarová prostředí.. Například dveře mohou být projektantem v modelu CAD označeny jako prvek kategorie "dveře", ale při přenosu do systému ERP nebo PMIS mohou být chybně označeny, ať už z důvodu chyby na straně projektanta, nebo z důvodu nekonzistence mezi systémy. Navíc tento prvek často ztrácí některé důležité atributy nebo mizí z účetnictví systému při exportech a importech dat. To vede k mezerám v toku dat a narušuje zásadu komplexní digitalizace stavebních procesů. Vzniká tak kritická mezera mezi "viditelným" a "srozumitelným" sémantickým významem, která narušuje integritu dat a výrazně komplikuje automatizaci procesů v celém životním cyklu stavebního projektu.

Řešení výzvy univerzální klasifikace stavebních prvků pomocí technologií velkých dat a strojového učení (obr. 10.1-6) bude katalyzátorem transformace celého odvětví - a možná i nečekaným objevem pro mnohé stavební firmy. Jednotný, učenlivý klasifikační systém bude základem pro škálovatelnou analytiku, digitální řízení a zavádění umělé inteligence do každodenní praxe stavebních organizací.

Společnost NVIDIA a další technologické špičky již poskytují řešení v jiných odvětvích, která dokáží automaticky kategorizovat a strukturovat obrovské množství textových a vizuálních informací.

Například model NeMo Curator společnosti NVIDIA [161] se specializuje na automatickou klasifikaci a kategorizaci dat do předem definovaných kategorií, což hraje klíčovou roli při optimalizaci potrubí zpracování informací pro jemné ladění a předtrénování generativních modelů umělé inteligence. Platforma Cosmos je trénována na reálných videích a 3D scénách [162] a poskytuje základ pro autonomní systémy a digitální dvojčata, která již vznikají v ekosystému NVIDIA. NVIDIA Omniverse, která se do roku 2025 stala předním nástrojem pro práci s formátem USD, univerzálním popisem scény, který by mohl časem nahradit formát IFC v procesech přenosu informací o návrhu. Spolu s Isaac Sim, simulátorem robotických procesů [163], představují řešení jako NeMo Curator, Cosmos a Omniverse novou úroveň automatizace: od čištění a filtrování dat až po generování trénovacích sad, modelování vlastností objektů a trénink robotů na stavbě. Všechny tyto nástroje jsou navíc zdarma a s otevřeným zdrojovým kódem, což výrazně snižuje překážky jejich přjetí v inženýrské a stavební praxi.

Automatická klasifikace dat na úrovni strukturovaných tabulek není tak obtížná, jak by se mohlo na první pohled zdát. Jak jsme si ukázali v předchozí kapitole (obr. 9.1-10), je možné chybějící nebo nesprávné hodnoty tříd doplnit na základě obdobných parametrů jiných prvků, pokud existují nahromaděná historická data. Pokud již byly prvky s podobnými charakteristikami správně klasifikovány v několika dokončených projektech, může systém s vysokou pravděpodobností navrhnut vhodnou hodnotu pro nový nebo neúplný prvek (obr. 10.1-6). Taková logika založená na zprůměrovaných hodnotách a analýze kontextu může být obzvláště účinná při hromadném zpracování tabulkových dat pocházejících z odhadů, specifikací nebo modelů CAD.

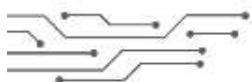


Obr. 10.1-6 Strojové učení vám může pomocí automaticky najít průměrné hodnoty pro nevyplněné (bílá pole) parametry tabulky na základě minulých projektů.

Na pozadí tak rychlého pokroku v oblasti strojového učení je jasné, že v roce 2025 je naivní se domnívat, že problém automatické klasifikace stavebních prvků zůstane dlouho nevyřešen. Ano, současné algoritmy ještě nedosáhly plné vyspělosti, zejména v případě neúplných nebo heterogenních dat, ale příležitost k přizpůsobení se rychle uzavírá.

Společnosti, které již investují do sběru, čištění a organizace svých dat a používají nástroje pro automatizaci ETL, budou ve výhodě. Ostatním hrozí, že zůstanou pozadu - stejně jako kdysi firmy nedokázaly čelit výzvám digitální transformace v dopravě a finančnictví

Ti, kteří se nadále spoléhají na manuální správu dat a tradiční metody odhadu nákladů a času, riskují, že se do začátku roku 2020 ocitnou v pozici vozových parků taxi z roku 2000, které se nedokážou přizpůsobit éře mobilních aplikací a automatizovaných výpočtů tras.



KAPITOLA 10.2.

PRAKTIKÝ PRŮVODCE ZAVÁDĚNÍM PŘÍSTUPU ZALOŽENÉHO NA DATECH

Od teorie k praxi: Plán digitální transformace ve stavebnictví

Stavebnictví postupně vstupuje do nové fáze vývoje, kdy jsou známé procesy stále více doplňovány - a někdy dokonce nahrazovány - digitálními platformami a transparentními modely interakce. To představuje pro společnosti nejen výzvy, ale také významné příležitosti. Ty organizace, které již dnes budují dlouhodobou digitální strategii, si budou moci nejen udržet svou pozici na trhu, ale také ji rozšířit tím, že zákazníkům nabídnu moderní přístupy a spolehlivá, technologicky podložená řešení.

Je důležité si uvědomit, že znalost pojmů a technologií je pouze výchozím bodem. Manažeři a odborníci stojí před praktickou otázkou: kde začít s implementací a jak proměnit teoretické myšlenky v reálnou hodnotu. Kromě toho se stále častěji objevuje otázka: na čem bude podnik založen, když tradiční metody kalkulace nákladů a časového rozvrhu může zákazník kdykoli přehodnotit.

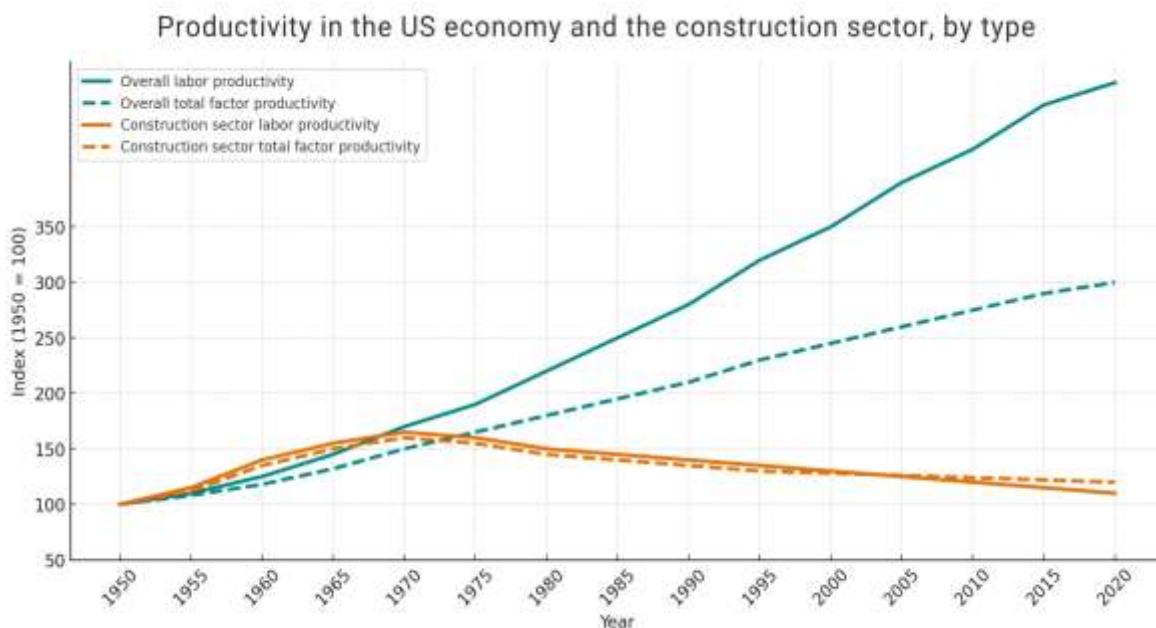
Odpověď pravděpodobně nespočívá ani tak v technologiích, ale ve vytvoření nové profesní kultury, v níž je práce s daty vnímána jako nedílná součást každodenní praxe. Právě nedostatečná pozornost věnovaná digitálním technologiím a inovacím naočkovala stavebnictví do vážného zaostávání, které lze pozorovat v posledních desetiletích [43].

Podle společnosti McKinsey činí výdaje na výzkum a vývoj ve stavebnictví méně než 1% příjmů, zatímco v automobilovém a leteckém průmyslu dosahují 3,5-4,5%. Podobně i náklady na IT ve stavebnictví zůstávají na úrovni nižší než 1% celkových příjmů [107].

V důsledku toho klesá nejen úroveň automatizace, ale i produktivita práce ve stavebnictví a v roce 2020 bude stavební dělník vyrábět méně než před půl stoletím (obr. 10.2-1)

Tyto problémy s produktivitou ve stavebnictví jsou společné pro většinu vyspělých i rozvojových zemí (produktivita ve stavebnictví klesla v 16 z 29 zemí OECD (obrázek 2.2-1)) a poukazují nejen na nedostatek technologií, ale také na potřebu systémových změn v samotném přístupu k řízení, vzdělávání a inovacím.

Úspěch digitální transformace nezávisí ani tak na počtu a dostupnosti nástrojů, ale na schopnosti organizací revidovat své procesy a rozvíjet kulturu otevřenou změnám. Klíčová není samotná technologie, ale lidé a procesy, které zajišťují její efektivní využívání, podporují neustálé učení a podporují přijímání nových myšlenek.



Obr. 10.2-1 Paradox produktivity práce a celkové produktivity zdrojů v ekonomice USA a ve stavebnictví (1950-2020) (podle [43]).

V prvních částech knihy byl model podnikatelského prostředí přirovnáván k lesnímu ekosystému (obrázek 2.1-2, obrázek 1.2-4, obrázek 1.3-2). Ve zdravém lese hrají periodické požáry, při vší své ničivé síle, klíčovou roli při dlouhodobé obnově. Vyčistí půdu od staré vegetace, vrátí do ní uložené živiny a vytvoří prostor pro nový život. Některé rostlinné druhy se dokonce vyvinuly tak, že se jejich semena otevří pouze tehdy, když jsou vystavena vysokým teplotám při požáru, což je přirozený mechanismus, který poskytuje ideální čas pro klíčení.

Podobně i v podnikání mohou krize působit jako "řízené vyhoření" a podpořit vznik nových přístupů a společností, které nejsou vázány na zastaralé systémy. Taková období nutí k opuštění neefektivních postupů a uvolňují zdroje pro inovace. Stejně jako les po požáru začíná s pionýrskými rostlinami, tak i podnik po krizi formuje nové, flexibilní procesy, které se stávají základem vyspělého informačního prostředí.

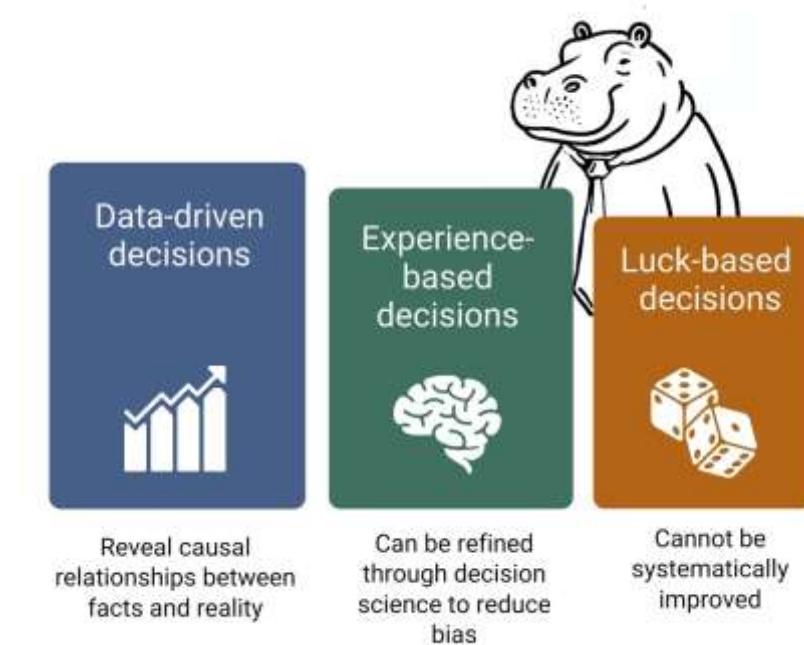
Společnosti, kterým se podaří správně interpretovat tyto "signální ohně" a přeměnit jejich rušivou energii v konstruktivní změnu, dosáhnou nové úrovně výkonnosti - s transparentnějšími a adaptivnějšími datovými procesy, které zvýší přirozenou schopnost organizace obnovovat se a růst.

O rostoucím vlivu umělé inteligence a strojového učení na podnikatelské prostředí již není pochyb. Nejedná se pouze o dočasný trend, ale o strategickou nutnost. Společnosti, které ignorují umělou inteligenci, riskují ztrátu konkurenční schopnosti na trhu, který stále více podporuje inovace a flexibilitu.

Budoucnost patří těm, kteří umělou inteligenci nevnímají jen jako nástroj, ale jako příležitost přehodnotit všechny aspekty svého podnikání - od optimalizace procesů až po manažerská rozhodnutí.

Položení digitálního základu: 1-5 kroků k digitální vyspělosti

V této kapitole se podíváme na plán digitální transformace a určíme klíčové kroky potřebné k zavedení přístupu založeného na datech, který může pomoci transformovat firemní kulturu i informační ekosystém společnosti.



Obrázek 10.2-2 Řízená aktualizace a výběr strategie: případ, zkušenost nebo data.

Podle studie společnosti McKinsey "Proč selhávají digitální strategie" (2018) existuje nejméně pět důvodů [164], proč se společnostem nedaří dosáhnout cílů digitální transformace

- **Nejasné definice:** Vedoucí pracovníci a manažeři chápou pojem "digitální" různě, což vede k nedorozuměním a rozporům.
- **Nepochopení digitální ekonomiky:** mnoho společností podceňuje rozsah změn, které digitalizace přináší do obchodních modelů a dynamiky odvětví (obrázek 10.1-6).
- **Ignorování ekosystémů:** společnosti se zaměřují na jednotlivá technologická řešení (datová sila) a přehlížejí potřebu integrace do širších digitálních ekosystémů (obrázek 2.2-2, obrázek 4.1-12).
- **Podcenění digitalizace ze strany konkurence:** manažeři neberou v úvahu, že digitální technologie aktivně zavádí i konkurenční výhody, což může vést ke ztrátě konkurenční výhody.
- **Chybějící dualita digitalizace:** generální ředitelé delegují odpovědnost za digitální transformaci na jiné vedoucí pracovníky, což byrokratizuje kontrolu a zpomaluje proces změn.

Řešení těchto problémů vyžaduje jasné pochopení a sladění digitálních strategií na všech úrovních organizace. Před vytvořením digitální strategie je důležité pochopit výchozí bod. Mnoho organizací má

tendenci přijímat nové nástroje a platformy, aniž by měly úplnou představu o současném stavu.

Krok 1: Provedte audit stávajících systémů a dat.

Před změnou procesů je důležité porozumět tomu, co už je zavedeno. Provedení auditu vám umožní identifikovat slabá místa ve správě dat a pochopit, jaké zdroje lze využít. Audit je jakýmsi "rentgenem" vašich podnikových procesů. Umožní vám identifikovat rizikové oblasti a určit, která data jsou pro váš projekt nebo podnikání kritická a která druhořadá.

Hlavní akce:

- Zmapujte své IT prostředí (v aplikaci Draw.io, Lucidchart, Miro, Visio nebo Canva). Vypište používané systémy (ERP, CAD, CAFM, CPM, SCM a další), které se podílejí na vašich procesech a o kterých jsme hovořili v kapitole "Technologie a systémy řízení v moderní výstavbě" (obr. 1.2-4)
- Posuďte problémy s kvalitou dat pro každý systém z hlediska četnosti duplicit, možných chybějících hodnot a nekonzistence formátu v každém systému.
- Identifikujte "bolestivá místa" - místa, kde se procesy mohou porouchat nebo často vyžadují manuální zásahy - importy, exporty a další validační procesy.

Pokud chcete, aby tým reportům důvěroval, musíte se od začátku ujistit, že jsou data správná.

Audit kvality dat vám ukáže, jaká data:

- Potřebuje další vývoj (je třeba nastavit automatické procesy čištění nebo další transformace).
- Jsou to "odpadky", které pouze upřávají systémy a kterých se lze zbavit tím, že se v procesech již nepoužívají.

Takový audit je možné provést na vlastní pěst. Někdy je však užitečné zapojit externího konzultanta - zejména z jiných oborů: nový pohled a nezávislost na stavebních "zvláštnostech" pomůže střízlivě posoudit současný stav a vyhnout se typickým nástrahám předpojatosti vůči určitým řešením a technologiím.

Krok 2: Určení klíčových standardů pro harmonizaci dat.

Po auditu je nutné vytvořit společná pravidla pro práci s daty. Jak jsme uvedli v kapitole "Standardy: Od náhodných souborů k inteligentnímu datovému modelu", pomůže to odstranit oddělené datové toky.

Bez jednotného standardu bude každý tým pracovat "po svém" a vy budete udržovat "zoo" integrací, kde se data ztrácejí při každé konverzi.

Hlavní akce:

■ Zvolte datové standardy pro výměnu informací mezi systémy:

- V případě tabulkových dat to mohou být strukturované formáty, jako je CSV, XLSX nebo efektivnější formáty, jako je Parquet.
- Pro výměnu volně strukturovaných dat a dokumentů: JSON nebo XML.

■ Zvládnout práci s datovými modely:

- Začněte parametrizací úloh na úrovni konceptuálního datového modelu - jak je popsáno v kapitole "Datové modelování: konceptuální, logický a fyzický model" (obr. 4.3-2).
- Jakmile proniknete hlouběji do logiky obchodního procesu, přejděte k formalizaci požadavků pomocí parametrů v logickém a fyzickém modelu (obrázek 4.3-6).
- Identifikujte klíčové entity, jejich atributy a vztahy v rámci procesů a vizualizujte tyto vztahy - jak mezi entitami, tak mezi parametry (obrázek 4.3-7).

■ K ověřování a standardizaci dat používejte regulární výrazy (RegEx) (obrázek 4.4-7), jak jsme o tom hovořili v kapitole "Strukturované požadavky a regulární výrazy RegEx". RegEx není složité, ale nesmírně důležité téma při práci na tvorbě požadavků na úrovni fyzických datových modelů.

Bez standardů na úrovni dat a vizualizace procesů není možné vytvořit konzistentní a škálovatelné digitální prostředí. Pamatujte, že "špatná data jsou drahá". A náklady na chyby se zvyšují s tím, jak se projekt nebo organizace stávají složitějšími. Sjednocení formátů, definování pojmenování, struktury a pravidel validace je investicí do stability a škálovatelnosti budoucích řešení.

Krok 3. Implementace DataOps a automatizace procesů.

Bez dobře definované architektury se budou společnosti nevyhnutelně potýkat s nesourodými daty obsaženými v oddělených informačních systémech. Data budou neintegrovaná, duplicitní na více místech a jejich údržba bude nákladná.

Představte si, že data jsou voda a datová architektura je složitý systém potrubí, kterým se voda dopravuje od zdroje uložení až k místu použití. Právě datová architektura určuje, jak jsou informace shromažďovány, ukládány, transformovány, analyzovány a dodávány koncovým uživatelům nebo aplikacím.

DataOps (Data Operations) je metodika, která integruje sběr, čištění, ověřování a používání dat do jediného automatizovaného procesního toku, jak jsme podrobně popsali v 8. části knihy.

Hlavní akce:

■ Vytvářet a přizpůsobovat ETL -konvejery pro automatizaci procesů:

- Výpis: organizujte automatický sběr dat z dokumentů PDF (obr. 4.1-2, obr. 4.1-5, obr. 4.1-7), tabulek Excel, modelů CAD (obr. 7.2-4), systémů ERP a dalších zdrojů, se kterými pracujete.

- Transformace: nastavení automatických procesů pro transformaci dat do jednotného strukturovaného formátu a automatizaci výpočtů, které budou probíhat mimo uzavřené aplikace (obr. 7.2-8).
- Načíst: zkuste vytvořit automatické nahrávání dat do souhrnných tabulek, dokumentů nebo centralizovaných úložišť (obr. 7.2-9, obr. 7.2-13, obr. 7.2-16).
- Automatizujte procesy výpočtu a QTO (Quantity Take-Off), jak jsme o tom hovořili v kapitole "QTO Quantity Take-Off: Seskupování dat projektu podle atributů":
 - Nastavení automatické extrakce objemů z modelů CAD - pomocí rozhraní API, zásuvných modulů nebo nástrojů pro reverzní inženýrství (obr. 5.2-5).
 - Vytvoření pravidel pro seskupování prvků pro různé třídy podle atributů ve formě tabulek (obrázek 5.2-12).
 - Zkuste automatizovat často se opakující výpočty objemu a nákladů mimo modulární uzavřené systémy (obr. 5.2-15).
- Začněte používat Python a Pandas ke zpracování dat, jak jsme si řekli v kapitole "Python Pandas: nepostradatelný nástroj pro práci s daty":
 - Použití nástroje DataFrame pro práci se soubory XLSX a automatizaci zpracování tabulkových dat (obr. 3.4-6).
 - Automatizace agregace a transformace informací pomocí různých knihoven Pythonu
 - Pomocí nástroje LLM můžete zjednodušit psaní hotových bloků kódu a celých potrubních linek (obrázek 7.2-18).
 - Zkuste si v jazyce Python vytvořit potrubní kanál, který najde chyby nebo zjistí anomálie a odešle oznámení odpovědné osobě (např. vedoucímu projektu) (obrázek 7.4-2).

Automatizace založená na principech DataOps umožňuje přejít od ručního a roztríštěného zpracování dat k udržitelným a opakovatelným procesům. To nejen snižuje zátěž zaměstnanců, kteří se denně zabývají stejnými transformacemi, ale také výrazně zvyšuje spolehlivost, škálovatelnost a transparentnost celého informačního systému.

Krok 4: Vytvoření otevřeného ekosystému správy dat.

Navzdory vývoji uzavřených modulárních systémů a jejich integraci s novými nástroji se společnosti potýkají s vážným problémem - rostoucí složitost těchto systémů převyšuje jejich užitečnost. Původní myšlenka vytvořit jedinou proprietární platformu pokrývající všechny podnikové procesy vedla k přílišné centralizaci, kdy jakékoli změny vyžadují značné zdroje a čas na přizpůsobení.

Jak jsme uvedli v kapitole "Firemní mycelium: Jak data propojují podnikové procesy", efektivní správa dat vyžaduje otevřený a jednotný ekosystém, který propojuje všechny zdroje informací.

Klíčové prvky ekosystému:

- Vyberte vhodné datové úložiště:

- Pro tabulky a výpočty použijte databáze - například PostgreSQL nebo MySQL (obr. 3.1-7).
- Pro dokumenty a zprávy může být vhodné cloudové úložiště (Google Drive, OneDrive) nebo systémy podporující formát JSON.
- Podívejte se na možnosti datových skladů, datových jezer a dalších nástrojů pro centralizované ukládání a analýzu velkého množství informací (obrázek 8.1-8).
- Implementace řešení pro přístup k vlastním datům:
 - Pokud používáte proprietární systémy, nakonfigurujte přístup k nim prostřednictvím API nebo SDK, abyste získali data pro externí zpracování (obr. 4.1-2).
 - Seznamte se s možnostmi nástrojů reverzního inženýrství pro formáty CAD (obrázek 4.1-13).
 - Nastavte ETL-pipeline, které pravidelně shromažďují data z aplikací nebo serverů, převádějí je do otevřených strukturovaných formátů a ukládají je do úložiště (obrázek 7.2-3).
 - Diskuse v týmu o tom, jak zajistit přístup k datům bez nutnosti používat proprietární software.
 - Pamatujte: data jsou důležitější než rozhraní. Dlouhodobou hodnotu přináší struktura a dostupnost informací, nikoli konkrétní nástroje uživatelského rozhraní.
- Zamyslete se nad vytvořením centra excelence (CoE) pro data, jak jsme o tom hovořili v kapitole "Centrum excelence (CoE) pro datové modelování", nebo nad tím, jak můžete poskytovat odborné znalosti v oblasti dat jiným způsobem (obrázek 4.3-9).

Ekosystém správy dat vytváří jednotný informační prostor, v němž všichni účastníci projektu pracují s konzistentními, aktuálními a ověřenými informacemi. Je základem pro škálovatelné, flexibilní a spolehlivé digitální procesy

Uvolnění potenciálu dat: 5-10 kroků k digitální zralosti

Kromě technické integrace je důležitým faktorem úspěšné implementace digitálních řešení jejich přijetí koncovými uživateli. Zapojení zákazníků nebo uživatelů do měření výkonnosti je výzvou jak pro zlepšení uživatelské zkušenosti, tak pro řízení změn ve společnosti. Pokud řešení nezypadá do známých pracovních postupů nebo neřeší skutečné problémy uživatelů či zákazníků, nebude používáno a žádná dodatečná opatření a pobídky to nespraví.

Transformace je iterativní proces založený na analýze dat o interakci uživatelů s novými procesy, s častými testovacími cykly, neustálou zpětnou vazbou a zdokonalováním.

Krok 5: Vytvoření kultury práce s daty, školení zaměstnanců a sběr zpětné vazby

Ani ten nejpokročilejší systém nebude fungovat bez zapojení zaměstnanců. Je třeba vytvořit prostředí, kde se data používají denně a kde tým chápe jejich hodnotu.

Ve zveřejněné zprávě britské vlády 2024 "Data Analytics and AI in Government Project Delivery" [83] se uvádí, že pro úspěšnou implementaci datové analytiky a AI je zásadní školení odborníků s potřebnými kompetencemi v oblasti zpracování a interpretace dat.

Nedostatek odborných znalostí v oblasti analýzy dat je jedním z klíčových problémů, které omezují digitální transformaci. Vedoucí pracovníci jsou zvyklí na zavedené postupy: čtvrtletní cykly, prioritní iniciativy a tradiční způsoby, jak posouvat projekty kupředu. Změna vyžaduje výrazného lídra - dostatečně vysoko postaveného, aby měl vliv, ale ne tak vysoko, aby měl čas a motivaci vést dlouhodobý transformační projekt.

Hlavní akce:

- Uvědomění si nutnosti přejít od subjektivního rozhodování založeného na názoru vysoce placeného zaměstnance (HiPPO) ke kultuře rozhodování založené na faktech a údajích, jak je uvedeno v kapitole "HiPPO aneb nebezpečí názoru v rozhodování" (obrázek 2.1-9).
- Organizujte systematické školení:
 - pořádat školení o tom, jak používat strukturovaná data, a zvát odborníky z jiných odvětví, kteří nejsou zaujati produkty a koncepty oblíbenými v dnešním stavebnictví.
 - Diskutovat s kolegy o přístupech a nástrojích pro analýzu dat a samostatně se učit praktickou práci s nástroji, jako je Python, pandas a LLM (obr. 4.1-3, obr. 4.1-6).
 - Vytvořit knihovnu výukových materiálů (nejlépe s krátkými videi) na téma strukturování dat (obr. 3.2-15) a vytváření datových modelů (obr. 4.3-6, obr. 4.3-7).
- Využívat moderní výukové technologie:
 - Použití jazykových modelů (LLM) pro podporu manipulace s kódem a daty, včetně generování kódu, refactoringu a analýzy, jakož i zpracování a interpretace tabulkových informací (obrázek 3.4-1).
 - Prozkoumejte, jak lze kód vytvořený pomocí LLM upravit a integrovat do kompletního řešení Pipeline při práci v offline vývojovém prostředí (IDE) (obr. 4.4-14, obr. 5.2-13).

Pokud manažer pokračuje v rozhodování "postaru", žádné školení nepřesvědčí lidi, aby brali analytiku vážně.

Budování datové kultury není možné bez neustálé zpětné vazby. Zpětná vazba umožňuje identifikovat nedostatky v procesech, nástrojích a strategiích, které nelze odhalit prostřednictvím interních reportů nebo formálních KPI metrik. Pochvalné komentáře uživatelů vašich řešení nepřenesou praktickou hodnotu. Cenná je až kritická zpětná vazba, zejména pokud je založena na konkrétních pozorováních a faktech. Získání takových informací však vyžaduje úsilí: musíte vytvořit procesy, v nichž mohou účastníci - interní i externí - sdílet připomínky (může mít smysl tak činit anonymně) bez zkreslení a bez obav, že by jejich názory mohly ovlivnit jejich vlastní práci. Je důležité, aby tak činili bez zkreslení a bez obav z negativních důsledků pro ně samotné.

Každé učení je nakonec samoučením [165].

- Milton Friedman, americký ekonom a statistik

Zavedení analytických nástrojů by mělo být doprovázeno pravidelným ověřováním jejich účinnosti v praxi (návratnost investic, klíčové ukazatele výkonnosti), čehož lze dosáhnout pouze prostřednictvím strukturované zpětné vazby od zaměstnanců, zákazníků a partnerů. To umožňuje společnostem nejen vyhnout se opakování chyb, ale také rychleji se přizpůsobit změnám v prostředí. Existence mechanismu pro sběr a analýzu zpětné vazby je jedním ze znaků vyspělosti organizace, která přechází od příležitostních digitálních iniciativ k udržitelnému modelu neustálého zlepšování (obrázek 2.2-5).

Krok 6: Od pilotních projektů k rozšíření

Vybírejte si dostatečně velké bitvy, aby na nich záleželo, a dostatečně malé, abyste je vyhráli.

- Jonathan Kozol

Zahájení digitální transformace "najednou a všude" je velmi riskantní. Efektivnějším přístupem je začít s pilotními projekty a postupně rozšiřovat úspěšné postupy.

Hlavní akce:

■ Zvolte správný projekt pro pilotní projekt:

- Definujte konkrétní obchodní cíl nebo proces s měřitelnými výsledky (KPI, ROI) (obr. 7.1-5).
- Zvolte proces automatizace ETL, například automatickou validaci dat nebo výpočet pracovní zátěže (QTO) pomocí jazyků Python a Pandas (obrázek 5.2-10)
- Stanovte jasné metriky úspěchu (např. zkrácení doby pro vypracování inspekčních specifikací nebo zpráv o validaci dat z týdne na jeden den).

■ Využívejte opakované přístupy:

- Začněte s jednoduchými procesy konverze dat a vytvořte proudové konverze víceformátových dat do formátů, které potřebujete pro své procesy (obr. 4.1-2, obr. 4.1-5).
- Postupně zvyšujte složitost úloh a rozšiřujte automatizaci procesů tím, že v IDE vytvoříte kompletní Potrubí založené na dokumentovaných blocích kódu (obr. 4.1-7, obr. 7.2-18).
- dokumentovat a zaznamenávat (nejlépe pomocí krátkých videí) úspěšná řešení a sdílet je s kolegy nebo v odborných komunitách.

- Vytvořit šablony a doprovodnou dokumentaci pro replikaci takových řešení, aby je mohli efektivně používat vaši kolegové (nebo členové odborné komunity, včetně uživatelů sociálních médií).

Postupný "roll-up" umožňuje udržet vysokou kvalitu změn a neupadnout do chaosu paralelních implementací. Strategie "od malého k velkému" minimalizuje rizika a umožňuje poučit se z malých chyb, aniž by přerostly v kritické problémy.

Přechod od projektového přístupu, kdy jsou zaměstnanci zapojeni jen částečně, k vytváření stálých týmů (např. odborných center - CoE) pomáhá zajistit udržitelný vývoj produktu i po vydání jeho první verze. Takové týmy nejen podporují stávající řešení, ale také je dále vylepšují.

Tím se snižuje závislost na zdlouhavém schvalování: členové týmu jsou oprávněni rozhodovat v rámci svých kompetencí. V důsledku toho jsou manažeři zbaveni potřeby mikromanagementu a týmy se mohou soustředit na vytváření skutečné hodnoty.

Vývoj nových řešení není sprint, ale maraton. Úspěšní jsou ti, kteří se zpočátku zaměřují na dlouhodobou a soustavnou práci.

Je důležité si uvědomit, že technologie vyžaduje neustálý vývoj. Investice do dlouhodobého vývoje technologických řešení jsou základem úspěšného provozu.

Krok 7: Používejte otevřené datové formáty a řešení

Jak jsme uvedli v kapitolách o modulárních platformách (ERP, PMIS, CAFM, CDE atd.), je důležité zaměřit se na otevřené a univerzální datové formáty, které zajistí nezávislost na řešeních dodavatelů a zvýší dostupnost informací pro všechny zúčastněné strany.

Hlavní akce:

- Přechod z uzavřených formátů na otevřené formáty:
 - Používejte otevřené formáty namísto proprietárních formátů nebo najděte způsob, jak nastavit automatické nahrávání nebo převod uzavřených formátů na otevřené (obrázek 3.2-15).
 - Implementace nástrojů pro práci s formáty Parquet, CSV, JSON, XLSX, což jsou výměnné standardy mezi většinou moderních systémů (obr. 8.1-2).
 - Pokud ve vašich procesech hraje důležitou roli práce s 3D geometrií, zvažte použití otevřených formátů, jako jsou USD, glTF, DAE nebo OBJ (obrázek 3.1-14).
- Použití vektorových databází databází k efektivní analýze a vyhledávání informací:
 - Použijte Bounding Box a další metody pro zjednodušení 3D -geometrie (obrázek 8.2-1).
 - Zamyslete se nad tím, kde můžete implementovat vektorizaci dat - převod textů, objektů nebo dokumentů na číselnou reprezentaci (obrázek 8.2-2).

■ Použití nástrojů pro analýzu velkých objemů dat:

- Organizovat ukládání nahromaděných historických dat (např. PDF, XLSX, CAD) do formátů vhodných pro analýzu (Apache Parquet, CSV, ORC) (obr. 8.1-2).
- Začněte používat základní statistické metody a pracovat s reprezentativními vzorky - nebo se alespoň seznamte se základními principy statistiky (obrázek 9.2-5).
- Implementovat a osvojit si nástroje pro vizualizaci dat a vztahů mezi daty a vizualizovat výsledky analýzy. Bez dobré vizualizace není možné plně pochopit samotná data ani procesy na nich založené (obrázek 7.1-4).

Přechod na otevřené datové formáty a zavedení nástrojů pro analýzu, ukládání a vizualizaci informací vytváří základ pro udržitelnou a nezávislou digitální správu. To nejen snižuje závislost na dodavatelích, ale také zajišťuje rovný přístup k datům pro všechny zúčastněné strany.

Krok 8. Začněte implementovat strojové učení pro predikci

Mnoho společností nashromáždilo obrovské množství dat - jakési "informační gejzíry", které jsou stále nevyužité. Tato data byla shromážděna v rámci stovek či tisíců projektů, ale v dalších procesech byla často využita pouze jednou nebo vůbec. Dokumenty a modely uložené v uzavřených formátech a systémech jsou často vnímány jako zastaralý a neužitečný balast. Ve skutečnosti jsou však cenným zdrojem - základem pro analýzu provedených chyb, automatizaci rutinních operací a vývoj inovativních řešení pro automatickou klasifikaci a rozpoznávání prvků v budoucích projektech.

Klíčovou výzvou je naučit se tato data získávat a transformovat je do použitelných poznatků. Jak je uvedeno v kapitole o strojovém učení a predikcích, techniky strojového učení mají potenciál výrazně zlepšit přesnost odhadů a predikcí v různých procesech souvisejících se stavebnictvím. Plné využití nashromážděných dat otevírá cestu ke zvýšení efektivity, snížení rizik a budování udržitelných digitálních procesů.

Hlavní akce:

■ Začněte s jednoduchými algoritmy:

- Zkuste použít lineární regresi - s využitím návodů z LLM - k předpovědi rekurentního výkonu v souborech dat, kde závislosti na velkém počtu faktorů chybí nebo jsou minimální (obrázek 9.3-4).
- Zvažte, ve kterých fázích vašich procesů by se teoreticky mohl uplatnit algoritmus k-nejbližších sousedů (k-NN) - například pro klasifikační úlohy, hodnocení podobnosti objektů nebo predikci na základě historických analogií (obr. 9.3-5).

■ Shromažďování a strukturování dat pro trénování modelů:

- Shromažďování historických údajů o projektu na jednom místě a v jednom formátu (obrázek 9.1-10).
- Práce na kvalitě a reprezentativnosti trénovacích vzorků prostřednictvím automatizovaného ETL (obrázek 9.2-8).

- Naučte se rozdělit data na trénovací a testovací množinu, jak jsme to udělali v příkladu s datovou sadou Titanic (obrázek 9.2-6, obrázek 9.2-7).
- Zvažte rozšíření použití technik strojového učení na širokou škálu úkolů, od předvídání časového harmonogramu projektu po optimalizaci logistiky, řízení zdrojů a včasné identifikaci potenciálních problémů.

Strojové učení je nástroj, který z archivovaných dat dělá cenný zdroj pro předvídání, optimalizaci a informované rozhodování. Začněte s malými soubory dat (obrázek 9.2-5) a jednoduchými modely a postupně zvyšujte jejich složitost.

Krok 9. Integrace IoT a moderních technologií sběru dat

Svět stavebnictví se rychle stává digitálním: každá stavební fotografie, každý příspěvek na Teams je již součástí širšího procesu parametrizace a tokenizace reality. Stejně jako kdysi GPS změnilo logistiku, mění internet věcí, RFID a technologie automatického sběru dat stavebnictví. Jak je uvedeno v kapitole "Internet věcí a chytré smlouvy", digitální staveniště se senzory a automatickým monitorováním je budoucností tohoto odvětví.

Hlavní akce:

- Implementace IoT -zařízení, RFID -značky a podrobné informace o procesech s nimi spojených:
 - vyhodnotit, ve kterých oblastech nebo fázích projektu lze dosáhnout největší návratnosti investic (ROI) - například při monitorování teploty, vibrací, vlhkosti nebo pohybu.
 - Zvažte použití RFID ke sledování materiálů, nástrojů a zařízení v celém dodavatelském řetězci.
 - Zvažte, jak lze shromážděná data integrovat do jednotného informačního systému, například Apache NiFi, pro automatizované zpracování a analýzu v reálném čase (obrázek 7.4-5).
- Zavedení systému monitorování v reálném čase:
 - Vytvářet informační panely pro sledování klíčových procesních nebo projektových ukazatelů pomocí vizualizačních nástrojů, jako jsou Streamlit, Flask nebo Power BI).
 - Nastavení automatických upozornění na kritické odchylky od plánu nebo norem (obr. 7.4-2).
 - Posoudit možnosti prediktivní údržby zařízení na základě shromážděných údajů a zjištěných vzorců (obrázek 9.3-6).
- Kombinujte data z různých zdrojů:
 - Začněte vizualizací datového modelu na fyzické úrovni - zohledněte strukturu informačních toků a klíčové parametry pocházející ze systémů CAD, zařízení IoT a platforem ERP (obr. 4.3-1).

- Začněte vytvořením hrubého náčrtu jednotné platformy pro analýzu dat a podporu rozhodování managementu. Zachyťte klíčové funkce, zdroje dat, uživatele a předpokládané scénáře použití (obrázek 4.3-7).

Čím dříve začnete propojovat reálné procesy s digitálním světem, tím dříve je budete moci řídit pomocí dat - efektivně, transparentně a v reálném čase.

Krok 10. Připravte se na budoucí změny v odvětví

Stavební společnosti jsou pod neustálým tlakem vnějšího prostředí: ekonomické krize, technologické skoky, změny předpisů. Stejně jako les, který musí odolávat dešti, sněhu, suchu i pálcímu slunci, žijí firmy v podmínkách neustálého přizpůsobování. A stejně jako se stromy stávají odolnými vůči mrazu a suchu díky hlubokému kořenovému systému, zůstávají životaschopné a konkurenceschopné pouze ty organizace, které mají pevný základ automatizovaných procesů, schopnost předvídat změny a pružně přizpůsobovat strategie.

Jak bylo uvedeno v kapitole "Strategie přežití: budování konkurenčních výhod", stavebnictví vstupuje do fáze radikální transformace. Interakce mezi objednatelem a zhovitelem se posouvá směrem k modelu přesvědčování, kde transparentnost, předvídatelnost a digitální nástroje nahrazují tradiční přístupy. V této nové realitě nevyhrávají ti největší, ale ti nejpružnější a technologicky nejvyspělejší.

Hlavní akce:

- Analýza zranitelných míst v podnikání v kontextu otevřených dat:
 - Zhodnotte, jak by demokratizace přístupu k datům v rámci ubrizgace mohla mít zničující dopad na vaši konkurenční výhodu a vaše podnikání (obrázek 10.1-5).
 - Promyslete strategii přechodu od neprůhledných a oddělených procesů k obchodním modelům založeným na otevřených řešeních, interoperabilitě systémů a transparentnosti dat (obr. 2.2-5).
- Vypracujte dlouhodobou digitální strategii:
 - Určete, zda chcete být lídrem v inovacích, nebo zda dáváte přednost scénáři "dohánění", při kterém budete šetřit své zdroje.
 - Popište fáze: krátkodobá (automatizace procesů, strukturování dat), střednědobá (implementace LLM a ETL), dlouhodobá (digitální ekosystémy, centralizovaná úložiště).
- Přemýšlejte o rozšíření portfolia svých služeb:
 - Zvážit nabídku nových služeb (zaměřených na energetickou účinnost, ESG, datové služby). O nových obchodních modelech budeme více hovořit v další kapitole.
 - Snažte se vystupovat jako spolehlivý technologický partner, který podporuje celý životní cyklus zařízení, od návrhu až po provoz. Důvěra ve vás by měla být založena na systematickém přístupu, transparentních procesech a schopnosti poskytovat udržitelná technologická řešení.

V transformačním prostředí nevyhrávají ti, kteří na změny pouze reagují, ale ti, kteří jsou proaktivní. Flexibilita, otevřenost a digitální vyspělost jsou základem udržitelnosti při budování zítřka.

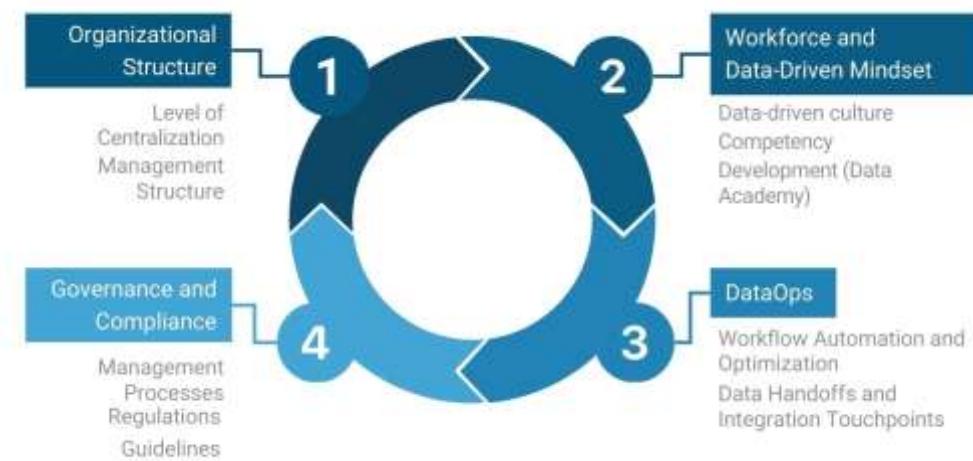
Plán transformace: od chaosu k datově řízené společnosti

Následující plán může sloužit jako výchozí měřítko pro vytvoření vlastní strategie digitální transformace založené na datech:

- **Audit a standardy:** analýza současného stavu, sjednocení dat
- **Strukturování a klasifikace dat:** automatizace transformace nestrukturovaných a volně strukturovaných dat.
- **Automatizace seskupení,** výpočtů a kalkulací: použití nástrojů a knihoven s otevřeným zdrojovým kódem pro automatizaci.
- **Ekosystém a COE:** vytvoření interního týmu pro vytvoření jednotného datového ekosystému ve společnosti.
- **Kultura a učení:** přechod od HiPPO -řešení k řešením založeným na datech
- **Pilotní projekty, zpětná vazba a rozšiřování:** postupujte iterativně: testujte nové metody v omezeném měřítku, sbírejte platnou zpětnou vazbu a postupně rozšiřujte řešení.
- **Otevřené formáty:** používejte univerzální a otevřené formáty, abyste byli nezávislí na dodavatelích softwaru.
- **Strojové učení:** začlenění algoritmů ML do procesů pro predikci a optimalizaci.
- **IoT a digitální staveniště:** integrace moderních technologií sběru dat do procesů
- **Strategické přizpůsobení:** příprava na budoucí změny v odvětví

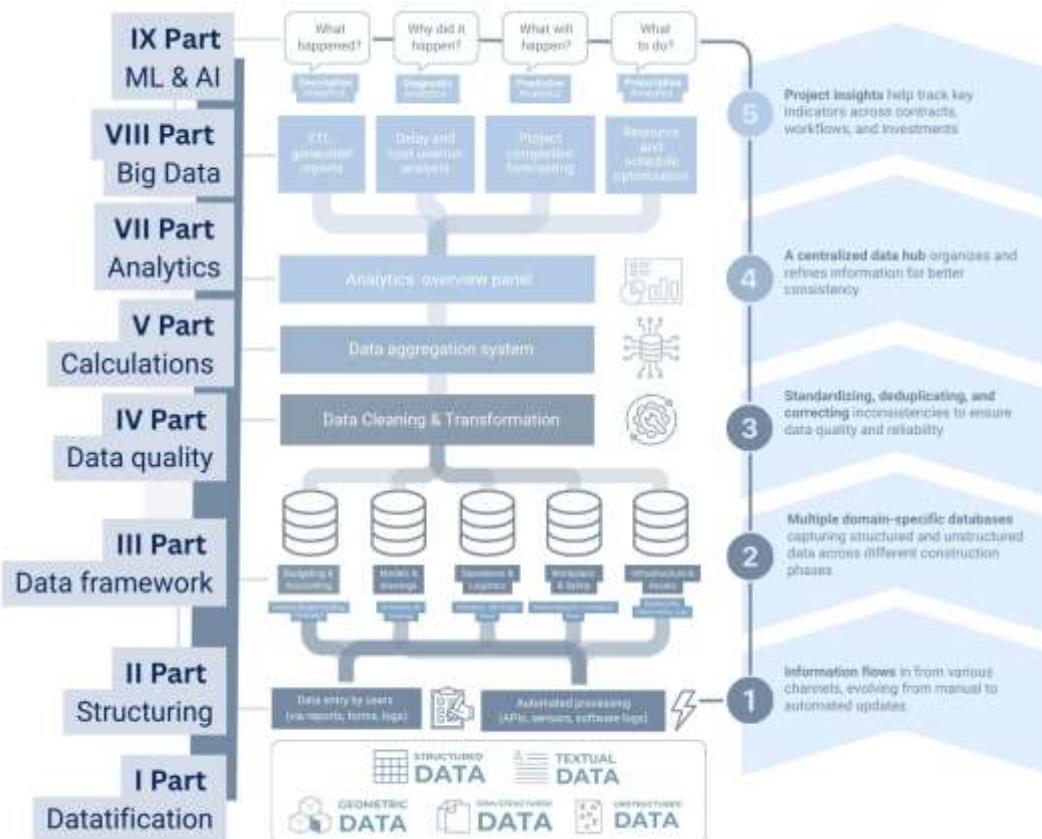
A hlavně nezapomeňte, že "samotná data firmu nezmění: mění ji lidé, kteří s nimi umí pracovat". Důraz na kulturu, transparentní procesy a závazek k neustálému zlepšování

Systémový přístup umožňuje přejít od oddělených digitálních iniciativ k plnohodnotnému modelu řízení založenému na datech, kde se rozhodnutí nezakládají na intuici nebo předpokladech, ale na datech, faktech a matematicky vypočítaných pravděpodobnostech. Digitální transformace stavebnictví nespočívá pouze v zavádění technologií, ale ve formování podnikového ekosystému, v němž se informace o projektech plynule a iterativně přenášejí mezi různými systémy. Algoritmy strojového učení přitom zajišťují automatickou průběžnou analýzu, prognózování a optimalizaci procesů. V takovém prostředí se spekulace a skrytá data stávají irrelevantními - zůstávají pouze ověřené modely, transparentní výpočty a předvídatelné výsledky.



Obr. 10.2-3 Klíčové prvky úspěšné správy dat na úrovni podniku.

Každá část knihy odpovídá určité fázi zpracování a analýzy dat ve stavebních projektech (obr. 2.2-5). Pokud se chcete vrátit k některému z dříve probíraných témat a podívat se na něj z hlediska uceleného chápání toku využití dat, můžete se podívat na názvy částí na obr. 10.2-4.



Obr. 10.2-4 Části knihy v kontextu pipeline zpracování dat (obr. 2.2-5): od digitalizace informací k analytice a umělé inteligenci.

Bez ohledu na velikost vaší organizace, úroveň technologické vyspělosti nebo rozpočet můžete začít přecházet na přístup založený na datech již dnes. I malé kroky správným směrem přinesou časem výsledky.

Transformace založená na datech není jednorázový projekt, ale nepřetržitý, opakující se proces zlepšování, který zahrnuje zavádění nových nástrojů, přepracování procesů a rozvoj kultury rozhodování založeného na datech.

Budování v průmyslu 5.0: Jak vydělávat peníze, když už se nemůžete skrývat

Stavební společnosti dlouho vydělávaly na neprůhlednosti procesů. Hlavním obchodním modelem byly spekulace - nadhodnocování nákladů na materiál, rozsah prací a procentuální přirážky v uzavřených systémech ERP - a PMIS - systémy, které jsou nepřístupné externím auditům. Omezený přístup zákazníků a jejich důvěryhodných osob k původním údajům o projektu vytvořil půdu pro schémata, v nichž bylo téměř nemožné ověřit spolehlivost výpočtů.

Tento model však rychle ztrácí na významu. S demokratizací přístupu k datům, vznikem LLM, příchodem otevřených dat a nástrojů pro automatizaci ETL přechází odvětví na nový standard práce.

Neprůhlednost tak přestává být konkurenční výhodou - brzy se stane břemenem, se kterým bude těžké se rozloučit. Transparentnost se změní z možnosti na podmínu pro udržení se na trhu.

S kým budou v nové digitální realitě spolupracovat zákazníci - banky, investoři, fyzičtí zákazníci, zákazníci soukromého kapitálu, zákazníci státní správy? Odpověď je zřejmá: s těmi, kteří jsou schopni poskytnout nejen výsledek, ale také zdůvodnění každého kroku na cestě k němu. S rostoucím objmem otevřených dat si partneři a zákazníci budou vybírat společnosti, které zaručují transparentnost, přesnost a předvídatelnost výsledků.

V této souvislosti vznikají nové obchodní modely, které jsou založeny na správě dat a důvěře, nikoli na spekulacích:

- **Prodej procesů namísto metrů čtverečních:** klíčovým přínosem se stává důvěra a efektivita, nikoliv konkrétní smlouvy s diskontem. Hlavní hodnotou bude předvídatelnost výsledku na základě spolehlivých a ověřených údajů. Moderní společnosti nebudou prodávat stavební objekt jako takový, ale:
 - přesné termíny a transparentní pracovní harmonogramy;
 - přiměřené odhady, podložené výpočty;
 - plná digitální sledovatelnost a kontrola ve všech fázích projektu.
- **Inženýrství a analytika jako služba:** model "Data jako služba" (způsob, jak uživatelům dodávat hotová data prostřednictvím internetu jako službu), kdy se každý projekt stává součástí digitálního datového řetězce a obchodní hodnota spočívá ve schopnosti tento řetězec spravovat. Společnosti se transformují na inteligentní platformy nabízející řešení založená na automatizaci a analytice:
 - Automatizovaná a transparentní příprava odhadů a plánů;
 - hodnocení rizik a načasování na základě algoritmů strojového učení;
 - výpočet environmentálních ukazatelů (ESG, CO₂, energetická účinnost);
 - generování zpráv z auditovaných otevřených zdrojů.
- **Produktivizace inženýrských znalostí:** vývoj společnosti může být opakován využíván v rámci

společnosti a distribuován jako samostatný produkt, což představuje další zdroj příjmů prostřednictvím digitálních služeb. V novém prostředí firmy vytvářejí nejen projekty, ale také digitální aktiva:

- knihovny komponent a šablon odhadů;
- automatizované ověřovací moduly;
- Open-source pluginy a skripty (prodej konzultací) pro práci s daty.

■ **Nový typ společnosti: Data-Driven Integrator:** hráč na trhu, který není závislý na konkrétních dodavatelích softwaru nebo modulárních systémech a není "uzamčen" v jediném softwarovém rozhraní. S daty operuje volně - a na nich staví svou konkurenceschopnost. Stavební firma budoucnosti není pouhým dodavatelem, ale informačním integrátorem, který je schopen pro zákazníka plnit následující funkce:

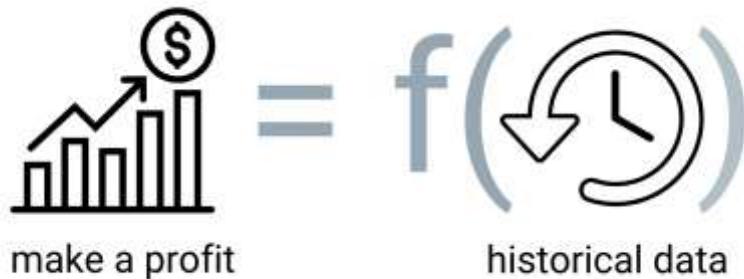
- Kombinujte data z různých zdrojů a provádějte analýzy;
- Zajištění transparentnosti a důvěryhodnosti procesů;
- poradenství v oblasti optimalizace obchodních procesů;
- vyvíjet nástroje, které pracují v ekosystému otevřených dat, LLM, ETL a Pipelines.

Průmysl 5.0 (obr. 2.1-12) znamená konec "éry ručního průměrování poměrů" a večerních schůzek generálních ředitelů s oddělením odhadu a účetnictví. Vše, co bylo dříve skryté - výpočty, odhady, objemy - se stává otevřeným, ověřitelným a srozumitelným i pro neoborníky. Vítězi budou ti, kteří se přeorientují jako první. Všichni ostatní zůstanou mimo novou digitální ekonomiku stavebnictví.

ZÁVĚR

Stavebnictví vstupuje do éry zásadních změn. Od prvních záznamů na hliněných tabulkách až po obrovské množství digitálních dat proudících z projektových serverů a stavenišť - historie správy informací ve stavebnictví vždy odrážela vyspělost technologie své doby. Dnes, s nástupem automatizace, otevřených formátů a inteligentních analytických systémů, nestojí toto odvětví před postupnou evolucí, ale před rychlou digitální transformací.

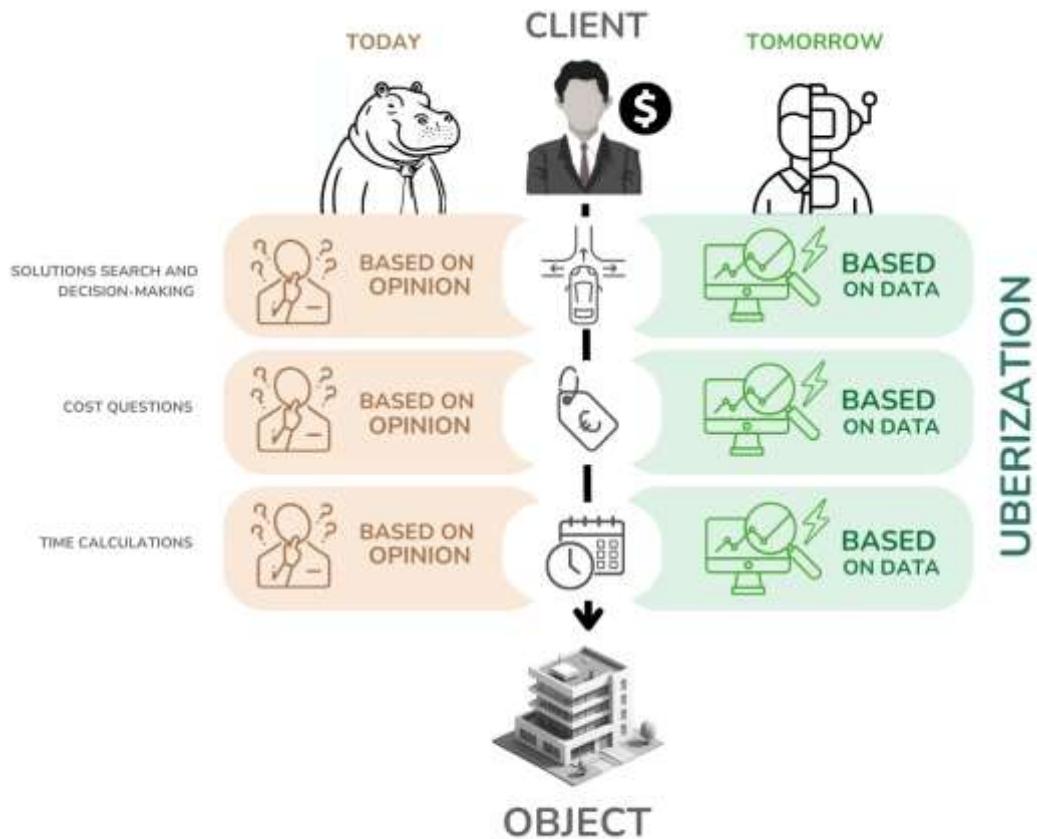
Stejně jako v jiných odvětvích hospodářství bude i ve stavebnictví nutné přehodnotit nejen nástroje, ale i principy práce. Společnosti, které dříve diktovaly trh a sloužily jako hlavní prostředník mezi klientem a projektem, ztrácejí své jedinečné postavení. Do popředí se dostává důvěra a schopnost pracovat s daty - od jejich sběru a strukturování až po analýzu, prognózu a automatizaci rozhodování.



Obr. 10.2-1 Strukturovaná historická data jsou palivem pro efektivní a řízený podnik.

Tato kniha podrobně popisuje klíčové principy správy dat ve stavebnictví - od auditu a standardizace po automatizaci procesů, využití vizualizačních nástrojů a implementaci inteligentních algoritmů. Podívali jsme se, jak lze i s omezenými zdroji vybudovat funkční datovou architekturu a začít se rozhodovat na základě ověřitelných faktů, nikoliv intuice. Práce s daty už není jen úkolem IT oddělení - stává se základem kultury řízení, která určuje flexibilitu, adaptabilitu a dlouhodobou udržitelnost firmy.

Aplikace technologií strojového učení, systémů automatického zpracování, digitálních dvojčat a otevřených formátů již dnes umožňuje eliminovat lidský faktor tam, kde byl dříve rozhodující. Výstavba směřuje k autonomii a kontrolovatelnosti, kdy pohyb od nápadu k realizaci projektu lze přirovnat k navigaci v režimu autopilotu: bez závislosti na subjektivních rozhodnutích, bez nutnosti manuálních zásahů v každé fázi, ale s plnou digitální sledovatelností a kontrolou (obr. 10.2-2).



Obr. 10.2-2 Posun od rozhodování na základě názorů významných odborníků (HiPPO) k analýze dat bude podporovat především zákazník.

Osvojíte-li si metody, principy a nástroje uvedené v této knize, budete moci ve své firmě začít přijímat rozhodnutí založená na datech, nikoli na intuici. Budete také schopni spouštět řetězce modulů v LLM, kopírovat hotové ETL Pipelines do svého vývojového prostředí (IDE) a automaticky zpracovávat data, abyste získali potřebné informace v požadované podobě. V návaznosti na kapitoly knihy věnované velkým datům a strojovému učení pak budete moci realizovat složitější scénáře - získávat nové poznatky z historických dat a aplikovat algoritmy strojového učení k předvídání a optimalizaci procesů.

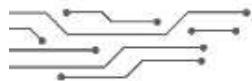
Otevřená data a procesy poskytnou základ pro přesnější odhady nákladů a harmonogramů projektů a zabrání stavebním společnostem spekulovat na základě neprůhledných údajů. Pro stavební průmysl je to výzva a zároveň příležitost přehodnotit svou roli a přizpůsobit se novému prostředí, kde se transparentnost a efektivita stanou klíčovými faktory úspěchu.

Klíčem k úspěchu v éře digitální transformace je ochota převzít znalosti a uplatnit je v praxi.

Společnosti, které si to uvědomí jako první, budou mít v nové digitální konkurenci výhodu. Je však důležité si uvědomit, že samotná data nic nezmění. Mnoho lidí bude muset změnit způsob myšlení, a to vyžaduje podněty. Vaše společnost musí přehodnotit způsob sdílení dat.

Lidé, kteří změní společnost, jsou ti, kteří umí s těmito daty pracovat, interpretovat je, využít je k optimalizaci a vytvořit na jejich základě novou architekturu procesů.

Pokud čtete tyto řádky, jste připraveni na změnu a jste již o krok napřed. Děkujeme vám, že jste si tuto cestu vybrali. Vítejte v éře digitální transformace!



O AUTOROVÍ

Jmenuji se Artyom Boiko. Moje cesta na stavbu začala v roce 2007 - prací horníka v dole na břidlicovou ropu v mé rodině městě a zároveň studiem na Vysoké škole báňské v Petrohradě, obor důlní a podzemní stavitelství. Na zadní straně obálky této knihy můžete vidět explosionalistu v čelbě, kde jsme těžili a odstřelovali stovky kubíků ropných břidlic. Moje kariéra se ubírala mnoha různými směry, od práce důlního a podzemního stavebního dělníka až po průmyslového horolezce, pokrývače a montéra výtahových zařízení. Měl jsem tu čest podílet se na projektech různého rozsahu, od výstavby soukromých domů až po velké průmyslové objekty v různých oblastech světa.



Postupem času se moje práce přesunula od fyzických konstrukcí ke správě informací a digitálním procesům. Od roku 2013 jsem pracoval na různých pozicích v malých, středních i velkých stavebních firmách v několika regionech Německa, od projektanta až po manažera správy dat. Co se týče správy dat, mé zkušenosti spočívají v práci s daty v různých ERP systémech, CAD (BIM), MEP, FEM, CMS. Zabýval jsem se optimalizací, automatizací procesů, ale i analýzou, strojovým učením, zpracováním dat ve fázích plánování, výpočtu a realizace stavebních prací v průmyslových, bytových, infrastrukturních a komunálních stavebních firmách.

S otevřeným softwarem a otevřenými daty pracuji od roku 2003. Za tu dobu jsem realizoval mnoho webových projektů - od webových stránek a internetových obchodů až po plnohodnotné webové aplikace využívající open source řešení a otevřené CMS. Tyto platformy, v mnoha ohledech podobné moderním stavebním ERP, mají modulární architekturu, vysokou přizpůsobivost a přístupnost. Tyto zkušenosti položily základ mého profesionálního přístupu - zaměření na open source technologie a kulturu společného vývoje. Respekt k open source a volné výměně znalostí se snažím ve stavebnictví prosazovat. Moje práce na zlepšení dostupnosti dat ve stavebnictví se promítla do vytvoření několika komunit na sociálních sítích, které diskutují o otevřenosti dat a využití otevřeného softwaru ve stavebnictví, a také do založení několika startupů vyvíjejících řešení pro zpřístupnění dat z různých uzavřených systémů a platform.

Můj přínos odborné komunitě spočívá v účasti jako řečníka na konferencích zabývajících se interoperabilitou CAD (BIM), ERP, 4D-5D, strojovým učením LLM a umělou inteligencí, jakož i v článcích publikovaných v evropských publikacích o stavebnictví. Jedním z mých významných úspěchů je vytvoření "Historie BIM". [111], komplexní mapy důležitých softwarových řešení pro správu dat ve stavebnictví. Moje sedmidílná série článků "Vývoj BIM a lobbistické hry", přeložená do několika jazyků, byla široce uznávána jako pokus o osvětlení skryté dynamiky digitálních standardů.

Tak jsem se od těžby kamene dostal k těžbě a systematizaci stavebních dat. Jsem vždy otevřený odbornému dialogu, novým nápadům a společným projektům. S vděčností příjmu jakoukoli zpětnou vazbu a těším se na vaše zprávy nebo na setkání s vámi mezi mými sledujícími na sociálních sítích. Moc vám děkuji, že jste tuto knihu dočetli až do konce! Budu rád, pokud vám tato kniha pomůže lépe porozumět tématu dat ve stavebnictví.

DALŠÍ VZTAHY

Názory čtenářů hrají důležitou roli při dalším vývoji publikací a výběru prioritních témat. Zvláště cenné jsou připomínky k tomu, které myšlenky se ukázaly jako užitečné a které vyvolaly pochybnosti a vyžadují další objasnění nebo uvedení zdrojů. Kniha obsahuje širokou škálu materiálů a analýz, z nichž některé se mohou zdát kontroverzní nebo subjektivní. Pokud v průběhu čtení narazíte na nepřesnosti, nesprávně citované zdroje, logické nesrovonalosti nebo překlepy, budu vám vděčný za vaše připomínky, myšlenky nebo kritiku, které můžete zaslat na adresu:boikoartem@gmail.com. Nebo prostřednictvím zpráv na LinkedIn:linkedin.com/in/boikoartem

Budu vděčný za jakoukoli zmínku o knize Data-Driven Construction na sociálních sítích - sdílení zážitků z četby pomáhá šířit informace o otevřených datech a nástrojích a podporuje mou práci.

KOMENTÁŘ K PŘEKLADU

Tato kniha byla přeložena pomocí technologie umělé inteligence. Díky tomu byl proces překladu mnohem rychlejší. Jako u každé technologické operace se však mohou vyskytnout chyby nebo nepřesnosti. Pokud si všimnete něčeho, co se vám zdá být nesprávné nebo nesprávně přeložené, napište mi prosím. Vaše připomínky pomohou zlepšit kvalitu překladu.

DATOVĚ ŘÍZENÉ STAVEBNÍ KOMUNITY

Je to místo, kde se můžete svobodně ptát a sdílet své problémy a řešení:

DataDrivenConstruction.io: <https://datadrivenconstruction.io>

LinkedIn: <https://www.linkedin.com/company/datadrivenconstruction/>

Twitter: <https://twitter.com/datadrivenconst>

Telegram: <https://t.me/datadrivenconstruction>

YouTube: [@datadrivenconstruction](https://www.youtube.com/@datadrivenconstruction)

DALŠÍ DOVEDNOSTI A POJMY

Kromě klíčových principů práce s daty ve stavebnictví se DataDrivenConstruction zabývá širokou škálou dalších konceptů, programů a dovedností, které jsou pro profesionály zaměřené na data nezbytné. Některé z nich jsou uvedeny pouze v přehledu, ale pro praxi jsou zásadní.

Zájemci mohou navštívit webové stránky DataDrivenConstruction.io, kde naleznou odkazy na další materiály týkající se klíčových dovedností. Tyto materiály zahrnují práci s jazykem Python a Pandas, vytváření ETL -procesů, příklady zpracování dat ve stavebních CAD projektech, systémy big data a moderní přístupy k vizualizaci a analýze stavebních dat.

Při přípravě knihy "DataDrivenConstruction" a všech případových studií bylo použito mnoho nástrojů a softwaru s otevřeným zdrojovým kódem. Autor by rád poděkoval vývojářům a spoluautorům těchto řešení:

- Python a Pandas - základ manipulace s daty a automatizace
- Scipy, NumPy, Matplotlib a Scikit-Learn - knihovny pro analýzu dat a strojové učení
- SQL a Apache Parquet - nástroje pro ukládání a zpracování velkého množství stavebních dat.
- Open Source CAD (BIM) nástroje pro otevřená data v otevřených formátech.
- N8n, Apache Airflow, Apache NiFi - systémy pro orchestraci a automatizaci pracovních postupů.
- DeepSeek, LlaMa, Mistral - Open Source LLM.

Zvláštní poděkování patří všem účastníkům diskusí na téma otevřených dat a nástrojů v profesních komunitách a sociálních sítích, jejichž kritika, připomínky a nápady pomohly zlepšit obsah a strukturu této knihy.

Sledujte vývoj projektu na webových stránkách DataDrivenConstruction.io, kde jsou zveřejňovány nejen aktualizace a opravy knihy, ale také nové kapitoly, výukové programy a praktické příklady použití popsaných technik.

MAXIMÁLNÍ POHODLÍ S TIŠTĚNOU VERZÍ

Držíte v ruce bezplatnou digitální verzi knihy **Data-Driven Construction**. Pro pohodlnější práci a rychlý přístup k materiálům doporučujeme věnovat pozornost **tištěnému vydání**:



■ **Vždy po ruce:** kniha v tištěné podobě se stane spolehlivým pracovním nástrojem, který vám umožní rychle najít a použít správné vizualizace a diagramy v jakékoli pracovní situaci.

■ **Vysoko kvalitní ilustrace:** všechny obrázky a grafika v tištěném vydání jsou prezentovány v maximální kvalitě.

■ **Rychlý přístup k informacím:** snadná navigace, možnost dělat si poznámky, záložky a pracovat s knihou kdekoli.

Zakoupením plné tištěné verze knihy získáte praktický nástroj pro pohodlnou a efektivní práci s informacemi: možnost rychle využívat obrazové materiály při každodenních úkolech, rychle vyhledávat potřebná schémata a dělat si poznámky. Navíc svým nákupem podpoříte šíření otevřených znalostí.

Tištěnou verzi knihy si můžete objednat na adrese: datadrivenconstruction.io/books



JEDINEČNÁ PŘÍLEŽITOST PRO STRATEGICKÉ UMÍSTĚNÍ

Nabízíme vám možnost umístit reklamní materiály do bezplatné verze DataDrivenConstruction. Placená verze publikace již první rok po svém vydání přitáhla pozornost odborníků z více než 50 zemí světa - od Latinské Ameriky po asijsko-pacifickou oblast. Chcete-li projednat individuální podmínky spolupráce a získat podrobné informace o možnostech umístění, vyplňte prosím formulář pro zpětnou vazbu na oficiálním portálu datadrivenconstruction.io nebo napište na kontakty uvedené na konci knihy.



KAPITOLY KNIHY JSOU K DISPOZICI NA ADRESE DATADRIVENCONSTRUCTION.IO

Kapitoly z knihy Data-Driven Construction si můžete přečíst na webových stránkách Data-Driven Construction, kde jsou postupně zveřejňovány části knihy, abyste mohli rychle najít potřebné informace a využít je při své práci. Najdete zde také mnoho dalších publikací na podobná témata a příklady aplikací a řešení, které vám pomohou rozvíjet vaše dovednosti a aplikovat data ve stavebnictví.



NEJNOVĚJŠÍ VERZE KNIHY STÁHNOUT Z OFICIÁLNÍCH WEBOVÝCH STRÁNEK

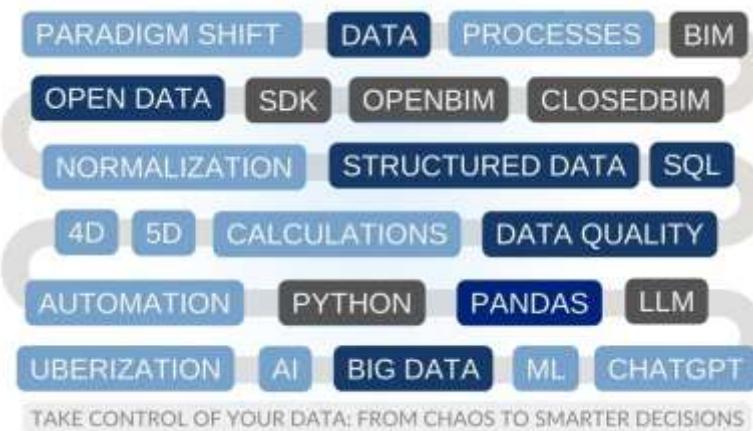
Aktuální a nejnovější verze knihy DataDrivenConstruction jsou k dispozici ke stažení na adrese datadrivenconstruction.io. Pokud chcete dostávat aktualizace s novými kapitolami knihy, praktickými tipy nebo recenzemi nových aplikací, přihlaste se k odběru newsletteru:

- Jako první se seznámíte s novými částmi knihy.
- Získejte praktické případové studie a rady týkající se analytiky a automatizace ve stavebnictví.
- Sledujte aktuální trendy, publikace a příklady použití.

Přejděte na datadrivenconstruction.io a přihlaste se k odběru!

KONSTRUKCE ZALOŽENÁ NA DATECH: PORADENSTVÍ, WORKSHOPY A ŠKOLENÍ

Školící programy a poradenství společnosti DataDrivenConstruction pomohly desítkám předních stavebních společností po celém světě zvýšit efektivitu, snížit náklady a zlepšit kvalitu řešení. Mezi klienty DataDrivenConstruction patří některí z největších hráčů na miliardovém trhu, včetně stavebních, poradenských a IT společností.



DataDrivenConstruction zaměřuje:

- **Správa kvality dat:** pomáhá parametrizovat úlohy, shromažďovat požadavky, ověřovat a připravovat data pro automatizované zpracování.
- **Data Mining - extrakce a strukturování dat:** nastavení procesů ETL a extrakce dat z e-mailů, PDF, Excel, obrázků a dalších zdrojů.
- **Analytika BIM a CAD :** shromažďování, strukturování a analýza informací ze souborů RVT, IFC, DWG a dalších formátů CAD (BIM).
- **Analytika a transformace dat:** přeměna různorodých informací na strukturovaná data, analytiku, náhledy a rozhodnutí.
- **Integrace dat a automatizace procesů:** od automatizovaného vytváření dokumentů po integraci s interními systémy a externími databázemi.

Kontaktujte DataDrivenConstruction.io a zjistěte, jak může automatizace pomoci vaší společnosti dosáhnout hmatatelných obchodních výsledků.

Proč si vybrat právě nás?

- **Relevance:** hovoří o hlavních trendech a poznatkách v oboru
- **Praktické:** pomáhá profesionálům efektivně řešit každodenní úkoly prostřednictvím PoC.
- **Individuální přístup:** šitý na míru vaší firmy, maximalizace přínosů školení a poradenství

Hlavní oblasti, na které se tým Da-

SLOVNÍČEK

AI (Artificial Intelligence) - umělá inteligence; schopnost počítačových systémů provádět úkoly, které obvykle vyžadují lidskou inteligenci, jako je rozpoznávání vzorů, učení a rozhodování.

Apache Airflow je open source platforma pro orchestraci pracovních postupů, která umožňuje programově vytvářet, plánovat a sledovat pracovní postupy a ETL pomocí DAG (směrovaných acyklických grafů).

Apache NiFi je nástroj pro automatizaci datových toků mezi systémy, který se specializuje na směrování a transformaci dat.

Apache Parquet je efektivní souborový formát pro sloupcové ukládání dat, optimalizovaný pro použití v systémech pro analýzu velkých objemů dat. Poskytuje výraznou kompresi a rychlé zpracování.

API (Application Programming Interface) - formalizované rozhraní, které umožňuje jednomu programu komunikovat s druhým bez přístupu ke zdrojovému kódu a vyměňovat si data a funkce prostřednictvím standardizovaných požadavků a odpovědí.

Atribut - charakteristika nebo vlastnost objektu, která popisuje jeho vlastnosti (např. plocha, objem, cena, materiál).

Databáze jsou organizované struktury pro ukládání, správu a přístup k informacím, které slouží k efektivnímu vyhledávání a zpracování dat.

BEP (BIM Execution Plan) - plán implementace informačního modelování budov, který definuje cíle, metody a procesy implementace BIM v projektu.

Big Data - pole informací o značném objemu, rozmanitosti a rychlosti aktualizace, která vyžadují speciální technologie pro zpracování a analýzu.

BI (Business Intelligence) - Business **Intelligence**; procesy, technologie a nástroje pro přeměnu dat na smysluplné informace pro rozhodování.

BIM (Building Information Modeling) - informační modelování budov; proces vytváření a správy digitálních reprezentací fyzických a funkčních vlastností stavebních projektů, zahrnující nejen 3D modely, ale také informace o vlastnostech, materiálech, čase a nákladech.

BlackBox/WhiteBox - přístupy k pochopení systému: v prvním případě je vnitřní logika skrytá, viditelné jsou pouze vstupy a výstupy; ve druhém případě je zpracování transparentní a dostupné pro analýzu.

Bounding Box je geometrická konstrukce, která popisuje hranice objektu v trojrozměrném prostoru pomocí minimálních a maximálních souřadnic X, Y a Z a vytváří kolem objektu "box".

BREP (Boundary Representation) je geometrická reprezentace objektů, která je definuje pomocí hranic ploch.

CAD (Computer-Aided Design) je systém počítačového navrhování, který se používá k vytváření, úpravám a analýze přesných výkresů a 3D modelů v architektuře, stavebnictví, strojírenství a dalších odvětvích.

CAFM (Computer-Aided Facility Management) je software pro správu majetku a infrastruktury, který zahrnuje plánování prostor, správu majetku, údržbu a sledování nákladů.

CDE (Common Data Environment) - centralizovaný digitální prostor pro správu, ukládání, sdílení a spolupráci s projektovými informacemi ve všech fázích životního cyklu zařízení.

Centrum excelence (CoE) je specializovaná struktura v rámci organizace, která je zodpovědná za rozvoj určité oblasti znalostí, vývoj standardů a osvědčených postupů, školení zaměstnanců a podporu zavádění inovací.

CoClass je moderní systém klasifikace stavebních prvků třetí generace.

Konceptuální datový model je vysokourovňové znázornění základních entit a jejich vztahů bez detailů atributů, které se používá v počátečních fázích návrhu databáze.

CRM (Customer Relationship Management) je systém řízení vztahů se zákazníky, který slouží k automatizaci prodejních a servisních procesů.

DAG (Directed Acyclic Graph) je směrovaný acyklický graf používaný v systémech pro orchestraci dat (Airflow, NiFi) k určení posloupnosti úloh a závislostí.

Dash je framework v jazyce Python pro vytváření interaktivních webových vizualizací dat.

Přístrojový panel - Přístrojový panel, který vizuálně prezentuje klíčové ukazatele výkonnosti a metriky v reálném čase.

Přístup zaměřený na data je metodika, která upřednostňuje data před aplikacemi nebo softwarovým kódem a činí z dat hlavní aktivum organizace.

Správa dat - soubor postupů, procesů a zásad, které zajišťují vhodné a efektivní využívání dat v organizaci, včetně kontroly přístupu, kvality a zabezpečení.

Datové jezero je úložiště určené k ukládání velkého množství nezpracovaných dat v původním formátu až do jejich použití.

Data Lakehouse je architektonický přístup, který kombinuje flexibilitu a škálovatelnost datových jezer (**Data Lake**) se spravovatelností a výkonností datových skladů (DWH).

Výstavba řízená daty je strategický přístup, při kterém je každá fáze životního cyklu stavby - od návrhu až po provoz - podporována automatizovanými, vzájemně propojenými systémy. Tento přístup poskytuje průběžné učení založené na faktech, snižuje nejistotu a umožňuje společnostem dosáhnout trvale vedoucího postavení v oboru.

Data-Driven Integrator - společnost, která se specializuje na kombinování dat z různých zdrojů a jejich analýzu za účelem manažerského rozhodování.

Přístup založený na datech - metodika, při níž jsou data považována za strategický přínos a rozhodnutí jsou přijímána na základě objektivní analýzy informací, nikoli subjektivních názorů.

Minimalismus dat - přístup k redukci dat na nejcennější a nejvýznamnější, který umožňuje zjednodušené zpracování a analýzu informací.

Datová bažina - rozptýlená masa nestrukturovaných dat, která vzniká při nekontrolovaném shromažďování a ukládání informací bez řádné organizace.

DataOps je metodika, která kombinuje principy DevOps, data a analytiku a zaměřuje se na zlepšení spolupráce, integrace a automatizace datových toků.

Digitalizace informací je proces převodu všech aspektů stavební činnosti do digitální podoby vhodné pro analýzu, interpretaci a automatizaci.

DataFrame - Dvourozměrná tabulková datová struktura v knihovně Pandas, kde řádky představují jednotlivé záznamy nebo objekty a sloupce jejich vlastnosti nebo atributy.

Deskriptivní analýza - Analýza historických dat s cílem pochopit, co se stalo v minulosti.

Diagnostická analýza - Analýza dat s cílem zjistit, proč se něco stalo.

Gantův diagram je nástroj pro plánování projektů, který znázorňuje úkoly jako vodorovné sloupce na časové ose a umožňuje vizualizovat pořadí a dobu trvání prací.

DWH (Data Warehouse) je centralizovaný systém datového skladu, který shromažďuje informace z různých zdrojů, strukturuje je a zpřístupňuje pro analýzy a reporting.

ESG (Environmental, Social, Governance) - soubor kritérií pro hodnocení environmentálních, sociálních a správních dopadů společnosti nebo projektu.

ELT (Extract, Load, Transform) je proces, při kterém jsou data nejprve extrahována ze zdrojů, načtena do úložiště a poté transformována pro analytické účely.

ETL (Extract, Transform, Load) je proces extrakce dat z různých zdrojů, jejich transformace do požadovaného formátu a načtení do cílového úložiště pro analýzu.

ER-diagram (Entity-Relationship) - vizuální diagram zobrazující entity, jejich atributy a vztahy mezi nimi, používaný při modelování dat.

ERP (Enterprise Resource Planning) je komplexní modulární systém plánování podnikových zdrojů, který se používá k řízení a optimalizaci různých aspektů stavebního procesu.

Funkce - ve strojovém učení nezávislé proměnné nebo atributy používané jako vstupy do modelu.

Fyzický datový model - podrobná reprezentace struktury databáze, včetně tabulek, sloupců, datových typů, klíčů a indexů, optimalizovaná pro konkrétní DBMS.

FPDF je knihovna jazyka Python pro vytváření dokumentů PDF.

Geometrické jádro je softwarová komponenta, která poskytuje základní algoritmy pro vytváření, úpravy a analýzu geometrických objektů v aplikacích CAD, BIM a dalších inženýrských aplikacích.

HiPPO (Highest Paid Person's Opinion) - přístup k rozhodování založený spíše na názoru nejlépe placené osoby v organizaci než na objektivních údajích.

IDE (Integrated Development Environment) - integrované vývojové prostředí, komplexní nástroj pro psaní, testování a ladění kódu (např. PyCharm, VS Code, Jupyter Notebook).

IDS (Information Delivery Specification) je specifikace poskytování informací, která definuje požadavky na data v různých fázích projektu.

IFC (Industry Foundation Classes) je formát pro výměnu dat BIM, který zajišťuje interoperabilitu mezi různými softwarovými řešeními.

Průmysl 5.0 je koncept průmyslového rozvoje, který kombinuje možnosti digitalizace, automatizace a umělé inteligence s lidským potenciálem a udržitelností životního prostředí.

Integrace dat je proces spojování dat z různých zdrojů do jednoho uceleného systému, který poskytuje jednotný pohled na informace.

Informační síla jsou izolované systémy pro ukládání dat, které nesdílejí informace s ostatními systémy, což vytváří překážky pro efektivní využívání dat.

IoT (Internet věcí) je koncept připojení fyzických objektů k internetu za účelem shromažďování, zpracování a přenosu dat.

k-NN (k-Nearest Neighbors) je algoritmus strojového učení, který klasifikuje objekty na základě podobnosti s nejbližšími sousedy v trénovacím vzorku.

Kaggle je platforma pro soutěže v analýze dat a strojovém učení.

Kalkulace - výpočet nákladů na stavební práce nebo postupy pro určitou měrnou jednotku (např. 1 m² sádrokartonové stěny, 1 m³ betonu).

KPI (Key Performance Indicators) jsou **klíčové** ukazatele výkonnosti, kvantifikovatelné metriky používané k hodnocení úspěšnosti společnosti nebo konkrétního projektu.

Štítky - ve strojovém učení cílové proměnné nebo atributy, které by měl model předpovídat.

Algoritmus učení - proces hledání nejlepší hypotézy v modelu odpovídající cílové funkci pomocí souboru trénovacích dat.

Lineární regrese - statistická metoda modelování vztahu mezi závislou proměnnou a jednou nebo více nezávislými proměnnými.

LLM (Large Language Model) - Large Language Model, umělá inteligence vycvičená k porozumění a generování textu z obrovského množství dat, schopná analyzovat kontext a psát programový kód.

LOD (Level of Detail/Development) - úroveň podrobnosti modelu, která určuje míru geometrické přesnosti a informačního obsahu.

Logický datový model je podrobný popis entit, atributů, klíčů a vztahů odrážející obchodní informace a pravidla, mezistupeň mezi konceptuálním a fyzickým modelem.

Strojové učení - třída technik umělé inteligence, která umožňuje počítačovým systémům učit se a předpovídat na základě dat bez explicitního programování.

Masterformat je klasifikační systém první generace, který se používá ke strukturování stavebních specifikací do sekcí a oborů.

MEP (Mechanical, Electrical, Plumbing) - stavební inženýrské systémy, které zahrnují mechanické, elektrické a vodovodní komponenty.

Mesh je síťová reprezentace 3D objektů, která se skládá z vrcholů, hran a ploch.

Model - Ve strojovém učení soubor různých hypotéz, z nichž jedna approximuje cílovou funkci, kterou je třeba předpovědět nebo approximovat.

Datové modelování je proces vytváření strukturované reprezentace dat a jejich vztahů pro implementaci v informačních systémech, včetně konceptuální, logické a fyzické úrovni.

n8n je nástroj s otevřeným zdrojovým kódem pro automatizaci pracovních postupů a integraci aplikací pomocí nízkokódového přístupu.

Normalizace - ve strojovém učení proces, při kterém se různá číselná data srovnají do společného měřítka, aby se usnadnilo jejich zpracování a analýza.

Reverzní inženýrství - proces zkoumání zařízení, fungování a technologie výroby předmětu analýzou jeho struktury, funkcí a fungování. V kontextu dat - extrakce informací z proprietárních formátů pro použití v otevřených systémech.

OCR (Optical Character Recognition) je technologie **optického** rozpoznávání **znaků**, která převádí textové obrazy (naskenované dokumenty, fotografie) do strojově čitelného textového formátu.

OmniClass je mezinárodní klasifikační standard druhé generace pro správu stavebních informací.

Ontologie - systém vzájemných vztahů pojmu, který formalizuje určitou oblast znalostí.

Open Source - model vývoje a distribuce softwaru s otevřeným zdrojovým kódem, který je k dispozici k volnému použití, studiu a úpravám.

Open BIM je koncept otevřeného BIM, který zahrnuje používání otevřených standardů a formátů pro výměnu dat mezi různými softwarovými řešeními.

Otevřené standardy - veřejně dostupné specifikace pro dosažení konkrétního úkolu, které umožňují interakci a výměnu dat mezi různými systémy.

Pandas je open source knihovna jazyka Python pro zpracování a analýzu dat, která poskytuje datové struktury DataFrame a Series pro efektivní zpracování tabulkových informací.

Paradigma otevřených dat je přístup ke zpracování dat, kdy jsou informace volně k dispozici pro použití, opakování použití a šíření kýmkoli.

Parametrická metoda je metoda odhadu stavebního projektu, která využívá statistické modely k odhadu hodnoty na základě parametrů projektu.

PIMS (Project Information Model) je digitální systém určený k organizaci, ukládání a sdílení všech informací o projektu.

Pipeline - posloupnost procesů zpracování dat, od extrakce a transformace až po analýzu a vizualizaci.

PMIS (Project Information Management System) je systém řízení projektů určený pro detailní řízení úkolů na úrovni jednotlivých stavebních projektů.

Prediktivní analýza je část analytiky, která využívá statistické metody a strojové učení k předpovídání budoucích výsledků na základě historických dat.

Preskriptivní analytika - část analytiky, která nejen předpovídá budoucí výsledky, ale také navrhuje optimální kroky k dosažení požadovaných výsledků.

Proprietární formáty jsou uzavřené datové formáty kontrolované konkrétní společností, které omezují možnost sdílení informací a zvyšují závislost na konkrétním softwaru.

QTO (Quantity Take-Off) je proces získávání množství prvků z projektové dokumentace za účelem výpočtu množství materiálů potřebných k realizaci projektu.

Systém řízení kvality - systém řízení kvality, který zajišťuje, že procesy a výsledky splňují stanovené požadavky.

RAG (Retrieval-Augmented Generation) je metoda, která kombinuje generativní schopnosti jazykových modelů s extrakcí relevantních informací z podnikových databází, čímž zvyšuje přesnost a relevanci odpovědí.

RDBMS (Relational Database Management System) je relační systém správy databází, který organzuje informace ve formě vzájemně propojených tabulek.

RegEx (regulární výrazy) je formalizovaný jazyk pro vyhledávání a zpracování řetězců, který umožňuje zadávat šablony pro kontrolu textových dat na splnění určitých kritérií.

Regresy je statistická metoda analýzy vztahu mezi proměnnými.

Výpočty CO₂ jsou metodou odhadu emisí oxidu uhličitého spojených s výrobou a používáním stavebních materiálů a procesů.

Metoda zdrojů je metoda odhadu založená na podrobné analýze všech zdrojů (materiálů, práce, vybavení) potřebných k provedení stavebních prací.

RFID (Radio Frequency Identification) je technologie automatické identifikace objektů pomocí rádiových signálů, která se používá ke sledování materiálů, strojů a osob.

ROI (Return on Investment) je ukazatel vyjadřující poměr mezi ziskem a investovanými prostředky, který se používá k posouzení efektivnosti investic.

SaaS (Software as a Service) je model softwaru jako služby, kdy jsou aplikace hostovány poskytovatelem a zpřístupněny uživatelům prostřednictvím internetu.

SCM (Supply Chain Management) - řízení dodavatelského řetězce, které zahrnuje koordinaci a optimalizaci všech procesů od nákupu materiálu až po dodávku hotových výrobků.

Datová sila jsou izolovaná úložiště informací v rámci organizace, která nejsou integrována s ostatními systémy, což ztěžuje sdílení dat a je neefektivní.

SQL (Structured Query Language) je strukturovaný dotazovací jazyk používaný pro práci s relačními databázemi.

SQLite je lehký, vestavěný, multiplatformní DBMS, který nevyžaduje samostatný server a podporuje základní funkce SQL, široce používaný v mobilních aplikacích a vestavěných systémech.

Strukturovaná data - informace uspořádané v určitém formátu s jasnou strukturou, například v relačních databázích nebo tabulkách.

Volně strukturovaná data - informace s částečnou organizací a flexibilní strukturou, jako je JSON nebo XML, kde různé prvky mohou obsahovat různé sady atributů.

Entita je konkrétní nebo abstraktní objekt reálného světa, který lze jednoznačně identifikovat, popsat a reprezentovat ve formě dat.

Supervised Learning - typ strojového učení, při kterém je algoritmus trénován na rozdělených datech, kde je pro každý příklad znám požadovaný výsledek.

Taxonomie je hierarchický klasifikační systém, který se používá k systematickému třídění prvků na základě společných znaků.

Datová sada Titanic je oblíbená datová sada pro trénování a testování modelů strojového učení.

Trénování - proces, při kterém algoritmus strojového učení analyzuje data za účelem identifikace vzorů a vytvoření modelu.

Transferové učení je technika strojového učení, při níž se model vycvičený pro jednu úlohu použije jako výchozí bod pro jinou úlohu.

Transformace - proces změny formátu, struktury nebo obsahu dat pro pozdější použití.

Požadavky na data jsou formalizovaná kritéria, která definují strukturu, formát, úplnost a kvalitu informací potřebných pro podporu podnikových procesů.

Uberizace stavebnictví je proces transformace tradičních obchodních modelů ve stavebnictví pod vlivem digitálních platform, které zajišťují přímou interakci mezi zákazníky a dodavateli bez zprostředkovatelů.

Uniclass je systém klasifikace stavebních prvků druhé a třetí generace, který se široce používá ve Velké Británii.

USD (Universal Scene Description) je datový formát vyvinutý pro počítačovou grafiku, ale díky své jednoduché struktuře a nezávislosti na geometrických jádrech našel uplatnění i v technických systémech.

Validace dat je proces kontroly informací podle stanovených kritérií a požadavků, aby byla zajištěna přesnost, úplnost a konzistentnost údajů.

Vektorová databáze - Specializovaný typ databáze, která ukládá data jako vícerozměrné vektory pro efektivní sémantické vyhledávání a porovnávání objektů.

Vektorová reprezentace (embedding) je metoda transformace dat do vícerozměrných číselných vektorů, která umožňuje strojovým algoritmům efektivně zpracovávat a analyzovat informace.

VectorOps je metodika zaměřená na zpracování, ukládání a analýzu vícerozměrných vektorových dat, která je důležitá zejména v oblastech, jako jsou digitální dvojčata a sémantické vyhledávání.

Vizualizace - grafické znázornění dat pro lepší vnímání a analýzu informací.

Abecední kategorizace termínů byla provedena podle jejich anglických názvů.

SEZNAM ODKAZŮ A ONLINE MATERIÁLŮ

- [1] Gartner, "IT Key Metrics Data 2017: Index of Published Documents and Metrics", 12. prosince 2016. [Online]. Dostupné na: <https://www.gartner.com/en/documents/3530919>. [Datum přístupu: 1. března 2025].
- [2] KPMG, "Známé výzvy - nové přístupy. Globální průzkum stavebnictví 2023," 1. ledna 2023. [Online]. Dostupné na: <https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf>. [Přístup: 5. března 2025].
- [3] F. R. Barnard, "Obrázek vydá za tisíc slov", 10 Mapi 1927. [Online]. Dostupné na: https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words. [Datum vystoupení: 15. března 2025].
- [4] M. Bastian, "CEO Microsoftu Satya Nadella říká, že prohlašovat se za AGI je "nesmyslné hackování benchmarků"," 21. února 2025. [Online]. Dostupné na: <https://the-decoder.com/microsoft-ceo-satya-nadella-says-self-claimingagi-is-nonsensical-benchmark-hacking/>. [Datum přístupu: 15. března 2025].
- [5] W. E. Fórum, "Fórum utváří budoucnost stavebnictví - krajina v proměnách:", 1. ledna 2016. [Online]. Dostupné na: https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction.pdf. [Datum přístupu: 2. března 2025].
- [6] C. D. Gillespie, "Clay: Earth's entanglement in the age of clay", 2024. [Online]. Dostupné na: <https://ufl.pb.unizin.org/imos/chapter/clay/>.
- [7] "Papyrus ze 3. století př. n. l. Jazyk je řečtina," 2024. [Online]. Available: <https://www.facebook.com/429710190886668/posts/595698270954525>.
- [8] "Monitoring: využití dostupných nástrojů", 1980. [Online]. Dostupné na: <https://pubmed.ncbi.nlm.nih.gov/10246720/>. [Datum vystoupení: 15. března 2025].
- [9] PWC, "Data driven What students need to succeed in a rapidly changing business world", 15. února 2015. [Online]. Dostupné na: <https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf>. [Datum přístupu: 15. března 2025].
- [10] Skanska USA, "Trendy na podzimním stavebním trhu", 2. listopadu 2023. [Online]. Available: <https://x.com/SkanskaUSA/status/1720167220817588714>.
- [11] "Oxford Essential Quotations (4. vydání)", Oxford University Press, 2016. [Online]. Available:

- <https://www.oxfordreference.com/display/10.1093/acref/9780191826719.001.0001/q-or-ed4-00006236>. [Datum přístupu: 1. března 2025].
- [12] "Citát: Sondergaard o analýze dat," [Online]. Dostupné na:
<https://www.causeweb.org/cause/resources/library/r2493>. [Datum přístupu: 15. března 2025].
- [13] "Jak globální zájem o umělou inteligenci zvyšuje trh se správou dat", 28. května 2024. [Online]. Dostupné na: <https://iot-analytics.com/how-global-ai-interest-is-boosting-data-management-market/>. [Datum přístupu: 15. března 2025].
- [14] И. McCue, "Historie ERP", 2024. [Online]. Dostupné na:
<https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml>.
- [15] erpscout, "Cena ERP: Kolik stojí systém ERP?", [Online]. Dostupné na:
<https://erpscout.de/en/erp-costs/>. [Datum přístupu: 15. března 2025].
- [16] softwarepath, "Co nám 1384 projektů ERP říká o výběru ERP (2022 ERP report)", 18. 1. 2022. [Online]. Dostupné na: <https://softwarepath.com/guides/erp-report>. [Datum přístupu: 15. března 2025].
- [17] Deloitte, "Data-Driven Management in Digital Capital Projects", 16. prosince 2016. [Online]. Dostupné na:
<https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Real%20Estate/us-engineering-construction-data-driven-management-digital-capital-projects.pdf>. [Datum oslovení: 1. března 2025].
- [18] McKinsey, "The data-driven enterprise of 2025", 28. ledna 2022. [Online]. Dostupné na:
<https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025>. [Datum reference: 22. května 2024].
- [19] Wikipedia, "Mooreův zákon," [Online]. Dostupné na:
https://en.wikipedia.org/wiki/Moore%27s_law. [Datum přístupu: 15. března 2025].
- [20] Accenture, "Building More Value With Capital Projects", 1. ledna 2020. [Online]. Dostupné na:
<https://www.accenture.com/content/dam/accenture/final/a-com-migration/r3-3/pdf/pdf-143/accsenture-industryx-building-value-capital-projects-highres.pdf>. [Datum přístupu: 3. března 2024].
- [21] B. Marr, "Kolik dat denně vytvoříme? Ohromující statistiky, které by si měl přečíst každý," 2018. [Online]. Dostupné na: <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read>.
- [22] "Kolik dat se denně vyprodukuje?", 2024. [Online]. Dostupné na:
<https://graduate.northeastern.edu/resources/how-much-data-produced-every-day/>.

- [23] T. Sullivan, "AI and the global 'datasphere': how much information will humanity have by 2025?", 2024. [Online]. Dostupné na: <https://www.datauniverseevent.com/en-us/blog/general/AI-and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By-2025.html>.
- [24] Statista, "Total number of printed books produced in various regions of Western Europe in each half century between 1454 and 1800," [Internet]. [Online]. Dostupné na: <https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/>. [Datum vytvoření odkazu: 1. března 2025].
- [25] "Příklady cen", 2024. [Online]. Dostupné na: <https://cloud.google.com/storage/pricing-examples>.
- [26] M. Ashare, "Enterprises outsourcing data storage as complexity rises", 10. května 2024. [Online]. Dostupné na: <https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/>. [Datum přístupu: 15. března 2025].
- [27] JETSOFTPRO, "SaaS is Dead? Vysvětlení šokující předpovědi generálního ředitele Microsoftu," 13. ledna 2025. [Online]. Dostupné na: <https://jetsoftpro.com/blog/saas-is-dead/>.
- [28] BG2 Pod, "Satya Nadella | BG2 w/ Bill Gurley & Brad Gerstner," 12. prosince 2024. [Online]. Dostupné na: https://www.youtube.com/watch?v=9NtsnzRFJ_o. [Datum přístupu: 15. března 2025].
- [29] GoodReads, "Tim Berners-Lee," [online]. Dostupné na: <https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than>. [Datum přístupu: 15. března 2025].
- [30] KPMG, "Cue Construction 4.0: Make-or-Break Time", 1. ledna 2023. [Online]. Dostupné na: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [Přístup: 5. března 2025].
- [31] I. Deininger, B. Koch, R. Bauknecht a M. Langhans, "Using digital models for decarbonisation of a production site: An example of connecting a building model, a production model and an energy model", 2024. [Online]. Dostupné na: https://www.researchgate.net/publication/374023998_Using_Digital_Models_to_Decarbonize_a_Production_Site_A_Case_Study_of_Connecting_the_Building_Model_Production_Model_and_Energy_Model.
- [32] McKinsey, "REINVENTING CONSTRUCTION: A ROUTE TO HIGHER PRODUCTIVITY", 1. února 2017. [Online]. Dostupné na: <https://www.mckinsey.com/~/media/mckinsey/business%20functions/operations/our%20insights/reinventing%20construction%20through%20a%20productivity%20revolution/mgi-reinventing-construction-a-route-to-higher-productivity-full-report.pdf>.

- [33] Pracovní skupina pro stavebnictví místopředsedovi vlády, "Přehodnocení stavebnictví", 1. října 2014. [Online]. Dostupné na: https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf.
- [34] Forbes, "Bez názoru jste jen další člověk s daty", 15. března 2016. [Online]. Dostupné na: <https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/>. [Datum přístupu: 15. března 2025].
- [35] Wikiquote, "Charles Babbage," [online]. Dostupné na: https://en.wikiquote.org/wiki/Charles_Babbage. [Datum přístupu: 15. března 2025].
- [36] SAP, "New Research Finds That Nearly Half of Executives Trust AI Over Themselves", 12. března 2025. [Online]. Dostupné na: <https://news.sap.com/2025/03/new-research-executive-trust-ai/>. [Datum přístupu: 15. března 2025].
- [37] Kanadská stavební asociace a KPMG v Kanadě, 2021, "Stavebnictví v digitálním světě", 1. května 2021. [Online]. Dostupné na: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [Datum přístupu: 5. března 2025].
- [38] ZCS, "Dekódování páté průmyslové revoluce". [Online]. Dostupné na: <https://www.pwc.in/decoding-the-fifth-industrial-revolution.html>. [Datum přístupu: 15. března 2025].
- [39] M. K, Private Rights and Public Problems: The Global Economics of, Peterson Inst. for Intern. Economics, 2012.
- [40] F. N. a. Y. Z. Harvard Business School: Manuel Hoffmann, "The Value of Open Source Software", 24. 1. 2024. [Online]. Dostupné na: <https://www.hbs.edu/faculty/Pages/item.aspx?num=65230>. [Datum přístupu: 15. března 2025].
- [41] Naval Centre for Cost Analysis Air Force Cost Analysis Agency, "Software Development Cost Estimating Handbook", 1. září 2008. [Online]. Dostupné na: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%20I%20rev%2010.pdf>.
- [42] McKinsey, "Zvyšování produktivity ve stavebnictví" [online]. Dostupné na: <https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity>. [Datum oslovení: 15. března 2025].
- [43] A. G. a. C. Syverson, "The Strange and Awful Path of Productivity in the US Construction Sector", 19. 1. 2023. [Online]. Dostupné na: <https://bfi.uchicago.edu/insight/research-summary/the-strange-and-awful-path-of-productivity-in-the-us-construction-sector/>. [Datum přístupu: 1. března 2025].

- [44] McKinsey, "Delivering on construction productivity is no longer optional", 9. srpna 2024. [Online]. Dostupné na: <https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional>. [Datum přístupu: 5. března 2025].
- [45] ING Group, "Lagging productivity in construction is driving up construction costs," 12. prosince 2022. [Online]. Dostupné na: <https://think.ing.com/articles/lagging-productivity-drives-up-building-costs-in-many-eu-countries/>. [Datum oslovení: 15. března 2025].
- [46] M. Berman, "Šokující předpověď generálního ředitele Microsoftu: 'Agenti nahradí veškerý software'", 19. prosince 2024. [Online]. Dostupné na: <https://www.youtube.com/watch?v=uGOLYz2pgr8>. [Datum přístupu: 15. března 2025].
- [47] Business Insider, "CEO společnosti Anthropic tvrdí, že za 3 až 6 měsíců bude umělá inteligence psát 90 % kódu, který měli na starosti vývojáři softwaru," 15. března 2025. [Online]. Dostupné na: <https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3>. [Datum přístupu: 30. března 2025].
- [48] Statista, "Popularity comparison of database management systems (DBMSs) worldwide as of June 2024, by category", červen 2024. [Online]. Dostupné na: <https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/>. [Datum oslovení: 15. března 2025].
- [49] DB-Engines, "DB-Engines Ranking," [Online]. Dostupné na: <https://db-engines.com/en/ranking>. [Datum odkazu: 15. března 2025].
- [50] "Průzkum Stack Overflow mezi vývojáři 2023", 2024. [Online]. Dostupné na: <https://survey.stackoverflow.co/2023/>.
- [51] "SQL", 2024. [Online]. Dostupné na: <https://en.wikipedia.org/wiki/SQL>.
- [52] "Strukturovaná a nestrukturovaná data: Jaký je mezi nimi rozdíl?", 2024. [Online]. Dostupné na: <https://www.ibm.com/blog/structured-vs-unstructured-data/>.
- [53] DataDrivenConstruction, "PDF SROVNÁNÍ FORMÁTŮ DAT PRO STAVEBNÍ PROJEKTY", 23. 4. 2024. [Online]. Dostupné na: <https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-PROJECTS-1.pdf>.
- [54] "Building Information Modeling Whitepaper site," 2003. [Online]. Dostupné na: <https://web.archive.org/web/20030711125527/http://usa.autodesk.com/adsk/servlet/item?id=2255342&siteID=123112>.
- [55] A. Boyko, "Lobbying wars and BIM development. Část 5: BlackRock je pánem všech technologií.

- Jak korporace ovládají otevřený zdrojový kód," 2024. [Online]. Dostupné na: <https://bigdataconstruction.com/autodesk-oracle-blackrock-open-source/>.
- [56] D. Ushakov, "Direct Modeling - Who and Why Needs It? A Review of Competitive Technologies", 14. 11. 2011. [Online]. Dostupné na: https://isicad.net/articles.php?article_num=14805. [Datum oslovení: 02 2025].
- [57] C. Eastman a A. Cthers, "Eastman, Charles; And Cthers", září 1974. [Online]. Dostupné na: <https://files.eric.ed.gov/fulltext/ED113833.pdf>. [Datum přístupu: 15. března 2025].
- [58] D. Ushakov, "Direct Modeling - Who and Why Needs It? A Review of Competitive Technologies", 11. listopadu 2011. [Online]. Dostupné na: https://isicad.net/articles.php?article_num=14805. [Datum odkazu: 15. března 2025].
- [59] D. Weisberg, "Historie CAD," 12. prosince 2022. [Online]. Dostupné na: https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhistorynet. [Datum odkazu: 15. března 2025].
- [60] ADSK, "Bílá kniha Informační modelování budov", 2002. [Online]. Dostupné na: https://web.archive.org/web/20060512180953/http://images.adsk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Datum reference: 15. března 2025].
- [61] ADSK, "Bílá kniha Informační modelování budov v praxi," [online]. Dostupné na: https://web.archive.org/web/20060512181000/http://images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf. [Datum přístupu: 15. března 2025].
- [62] A. Boyko, "Lobbying wars and BIM development. Část 2: otevřený BIM VS uzavřený BIM. Evropa VS zbytek světa," 2024. [Online]. Dostupné na: <https://bigdataconstruction.com/lobbyist-wars-and-the-development-of-bim-part-2-open-bim-vs-closed-bim-revit-vs-archicad-and-europe-vs-the-rest-of-the-world/>.
- [63] A. Boyko, "Lobbykriege um Daten im Bauwesen | Techno-Feudalismus und die Geschichte von BIMs," 2024. [Online]. Dostupné na: https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360.
- [64] ADSK, "Whitepaper BIM", 2002. [Online]. Dostupné na: https://web.archive.org/web/20060512180953/http://images.autodesk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand. [Datum oslovení: 15. března 2025].
- [65] ADSK, "Integrated Design-Through-Manufacturing: Benefits and Rationale," [Online]. Dostupné na: https://web.archive.org/web/20010615093351/http://www3.adsk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Datum přístupu: 15. března 2025].

- [66] M. Shacklett, "Structured and unstructured data: Klíčové rozdíly," 2024. [Online]. Dostupné na: <https://www.datamation.com/big-data/structured-vs-unstructured-data/>.
- [67] K. Woolard, "Making sense of the growth of unstructured data", 2024. [Online]. Dostupné na: <https://automationhero.ai/blog/making-sense-of-the-rise-of-unstructured-data/>.
- [68] A. C. O. J. L. D. J. a. L. T. G. Michael P. Gallaher, "Cost Analysis of Inadequate Interoperability in the," 2004. [Online]. Dostupné na: <https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf> . [Datum přístupu: 02 2025].
- [69] CrowdFlower, "Data Science Report 2016", 2016. [Online]. Dostupné na: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf. [Datum přístupu: 15. března 2025].
- [70] Analyticsindiamag, "6 nejvíce časově náročných úkolů pro datové vědce", 15. května 2019. [Online]. Dostupné na: <https://analyticsindiamag.com/ai-trends/6-tasks-data-scientists-spend-the-most-time-doing/>.
- [71] BizReport, "Report: Data scientists spend bulk of time cleaning up", 6. července 2015. [Online]. Dostupné na: <https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-data-scientists-spend-bulk-of-time-cleaning-up.html>. [Datum přístupu: 5. března 2025].
- [72] S. Hawking, "Science AMA Series: Stephen Hawking AMA Answers!", 27. července 2015. [Online]. Dostupné na: https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_a_ma_answers/. [Datum přístupu: 15. března 2025].
- [73] B. Cyphers a K. Doctorow, "Privacy without Monopoly: Data Protection and Interoperability", 2024. [Online]. Dostupné na: <https://www.eff.org/wp/interoperability-and-privacy>.
- [74] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information", 1. října 2013. [Online]. Dostupné na: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>. [Datum přístupu: 15. března 2025].
- [75] A. Bojko, "Boj o otevřená data ve stavebnictví. Historie AUTOLISPu, intelliCADu, openDWG, ODA a openCASCADE," 15 05 2024. [Online]. Dostupné na: <https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry-2b97200e6393>. [Datum oslovení: 16. 02. 2025].
- [76] Wikipedia, "Microsoft a open source," [online]. Dostupné na: https://en.wikipedia.org/wiki/Microsoft_and_open_source. [Datum přístupu: 15. března 2025].

- [77] TIME, "The Gap Between Open and Closed AI Models Might Be Shrinking. Here's Why That Matters," 5. listopadu 2024. [Online]. Dostupné na: <https://time.com/7171962/open-closed-ai-models-epoch/>. [Datum přístupu: 15. března 2025].
- [78] The Verge, "Více než čtvrtina nového kódu ve společnosti Google je generována umělou inteligencí", 29. října 2024. [Online]. Available: <https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024>. [Datum přístupu: 15. března 2025].
- [79] McKinsey Digital, "The business case for using GPUs to accelerate analytics processing", 15. prosince 2020. [Online]. Dostupné: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics-zpracovani>. [Přístup: 15. března 2025].
- [80] PwC, "PwC Open Source Monitor 2019", 2019. [Online]. Dostupné na: <https://www.pwc.de/de/digitale-transformation/open-source-monitor-research-report-2019.pdf>. [Přístup 15. března 2025].
- [81] Travers Smith, "The Open Secret: Open Source Software", 2024. [Online]. Dostupné na: <https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source-software/>. [Datum přístupu: 15. března 2025].
- [82] Deloitte, "Proces přenosu dat při transformaci podniků", 2021. [Online]. Dostupné na: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-process-in-corporate-transformations.pdf>. [Datum reference: 15. března 2025].
- [83] gov.uk, "Data Analytics and AI in Government Project Delivery", 20. března 2024. [Online]. Dostupné na: <https://www.gov.uk/government/publications/data-analytics-and-ai-in-government-project-delivery/data-analytics-and-ai-in-government-project-delivery>. [Datum přístupu: 5. března 2025].
- [84] "Citát Původ: Všechno by mělo být co nejjednodušší, ale ne jednodušší," 13. května 2011. [Online]. Dostupné z: <https://quoteinvestigator.com/2011/05/13/einstein-simple/>. [Datum přístupu: 15. března 2025].
- [85] "Transformátor (architektura hlubokého učení)," [Online]. Dostupné na: [https://en.wikipedia.org/wiki/Transformer_\(deep_learning_architecture\)](https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture)). [Datum přístupu: 15. března 2025].
- [86] "Statistiky stahování balíčků Pythonu", 2024. [Online]. Dostupné na: <https://www.pepy.tech/projects/pandas>.
- [87] Rozhovor Bit, "Top 10 knihoven Pythonu", 2023. [Online]. Dostupné na: <https://www.interviewbit.com/blog/python->

- libraries/#:~:text=With%20more%20than%20137%2C000%20libraries,data%20manipulation%2C%20and%20many%20more. [Datum přístupu: 30. března 2025].
- [88] "NVIDIA a HP rozšiřují možnosti datové vědy a generativní umělé inteligence na pracovních stanicích," 7. března 2025. [Online]. Dostupné na: <https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science-generative-ai-workstations>. [Datum přístupu: 15. března 2025].
- [89] P. Orac, "Jak zpracovat DataFrame s miliony řádků za několik sekund", 2024. [Online]. Dostupné na: <https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds>.
- [90] Ç. Uslu, "Co je to Kaggle?", 2024. [Online]. Dostupné na: <https://www.datacamp.com/blog/what-is-kaggle>.
- [91] "Přednáška generálního ředitele společnosti NVIDIA Jensa Huanga na veletrhu COMPUTEX 2024," 2. června 2024. [Online]. Dostupné na: <https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w>. [Přístup 15. března 2025].
- [92] "Členové: Zakladatelé a firemní členové", 2024. [Online]. Dostupné na: <https://www.opendesign.com/member-showcase>.
- [93] A. Boyko, "Doba změn: IFC je minulostí aneb proč se ADSK a další dodavatelé CAD chtějí vzdát IFC za USD ve 14 klíčových faktech", 24. listopadu 2024. [Online]. Dostupné na: <https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82cccd10a>. [Přístup: 23. února 2025].
- [94] A. Bojko, "Svět po BIM. Přechod na data a procesy a zda stavebnictví potřebuje sémantiku, formáty a interoperabilitu," 20. 12. 2024. [Online]. Dostupné na: <https://boikoartem.medium.com/the-post-bim-world-7e35b7271119>. [Datum vystoupení: 23. února 2025].
- [95] N. I. o. Health, "NIH STRATEGIC PLAN FOR DATA SCIENCE", 2016. [Online]. Dostupné na: https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf. [Datum přístupu: 23. února 2025].
- [96] Harvard Business Review, "Bad Data Costs the U.S. \$3 Trillion Per Year", 22. září 2016. [Online]. Dostupné na: <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>.
- [97] Delpha, "Dopady kvality dat", 1. ledna 2025. [Online]. Dostupné na: <https://delpha.io/impacts-of-data-quality/>.
- [98] W. B. D. Guide, "Design for Maintainability: The Importance of Operations and Maintenance Considerations During the Design Phase of Construction Projects" (Návrh pro udržovatelnost:

- Význam úvah o provozu a údržbě ve fázi návrhu stavebních projektů). [Online]. Dostupné na: <https://www.wbdg.org/resources/design-for-maintainability>. [Datum přístupu: 15. března 2025].
- [99] O. o. D. C. P. a. Oversight, "Příručka pro plánování prevence a kontroly koroze vojenských systémů a zařízení", duben 2014. [Online]. Dostupné na: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guide%204%20Feb%202014.pdf>. [Datum přístupu: 15. března 2025].
- [100] Gartner, "Data Quality: Best Practices for Accurate Insights", 1. ledna 2025. [Online]. Dostupné na: <https://www.gartner.com/en/data-analytics/topics/data-quality>.
- [101] "For Want of a Nail," [Online]. Dostupné na: https://en.wikipedia.org/wiki/For_Want_of_a_Nail. [Datum přístupu: 15. března 2025].
- [102] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information", říjen 2013. [Online]. Dostupné: https://www.mckinsey.com/~media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/open%20data%20unlocking%20innovation%20and%20performance%20with%20liquid%20information/mgi_open_data_fullreport_oct2013.pdf. [Datum přístupu: 15. března 2025].
- [103] EY, "Cesta k uhlíkové neutralitě", 10. března 2023. [Online]. Dostupné na: https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality. [Datum přístupu: 15. března 2025].
- [104] PWC, "ESG Awareness", 1. července 2024. [Online]. Dostupné na: <https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf>. [Datum přístupu: 15. března 2025].
- [105] G. Hammond, "Embodied Carbon - The Inventory of Carbon and Energy (ICE)", 2024. [Online]. Dostupné na: <https://greenbuildingencyclopedia.uk/wp-content/uploads/2014/07/Full-BSRIA-ICE-guide.pdf>.
- [106] "CO₂_výpočet ztělesněného uhlíku", 2024. [Online]. Dostupné na: https://github.com/datadrivenconstruction/CO2_calculating-the-embodyied-carbon.
- [107] McKinsey, "Imagining Construction's Digital Future", 24. června 2016. [Online]. Dostupné na: <https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future>. [Datum přístupu: 25. února 2025].
- [108] Bund der Steuerzahler Deutschland e.V., "Das Schwarzbuch", 10. října 2024. [Online]. Dostupné na: <https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/>. [Datum oslovení: 15. března 2025].

- [109] SAS, "Datové jezero a datový sklad - poznejte rozdíl," [online]. Dostupné na: https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html. [Datum reference: 15. března 2025].
- [110] ADSK, "Informační modelování budov", 2002. [Online]. Dostupné na: https://www.laiserin.com/features/bim/autodesk_bim.pdf. [Datum oslovení: 15. března 2025].
- [111] A. Bojko, "Mapa historie BIM", 2024. [Online]. Dostupné na: <https://bigdataconstruction.com/history-of-bim/>.
- [112] A. S. Borkowski, "Definice BIM podle organizací a norem", 27. prosince 2023. [Online]. Dostupné na: <https://encyclopedia.pub/entry/53149>. [Datum oslovení: 5. března 2025].
- [113] CAD Vendor, "OPEN BIM Programme", 2012. [Online]. Dostupné na: https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/. [Datum reference: 30. března 2025].
- [114] Wikipedia, "Industry Foundation Classes," [Online]. Dostupné na: https://en.wikipedia.org/wiki/Industry_Foundation_Classes. [Datum přístupu: 15. března 2025].
- [115] Wikipedia, "IGES," [Online]. Dostupné na: <https://en.wikipedia.org/wiki/IGES>. [Datum odkazu: 30. března 2025].
- [116] A. Bojko, "Historie CAD (BIM)", 15. prosince 2021. [Online]. Dostupné na: https://miro.com/app/board/o9J_laML2cs=/. [Datum oslovení: 24. února 2025].
- [117] T. K. K. A. O. F. B. C. E. L. H. H. E. L. P. N. S. H. T. J. v. L. H. H. G. D. H. T. K. C. L. A. W. J. S. Francesca Noardo, "Reference study of IFC software support: the GeoBIM benchmark 2019 - Part I", 8. 1. 2021. [Online]. Dostupné na: <https://arxiv.org/pdf/2007.10951.pdf>. [Datum přístupu: 5. března 2025].
- [118] И. Rogachev, "Let's Talk BIM: Maxim Nechiporenko | Renga | IFC | Domestic BIM," 13. dubna 2021. [Online]. Dostupné na: <https://www.youtube.com/watch?t=3000&v=VO3Y9uuzF9M&feature=youtu.be>. [Datum oslovení: 5. března 2025].
- [119] D. Ares, "RETS in Real Estate: Why It's Crucial for Efficiency & Growth", 17. prosince 2024. [Online]. Dostupné na: <https://www.realpha.com/blog/rets-importance-in-real-estate-explained>. [Datum přístupu: 5. března 2025].
- [120] "Flex token cost", 2024. [Online]. Dostupné na: <https://www.adsk.com/buying/flex?term=1-YEAR&tab=flex>.
- [121] A. Boyko, "Forget BIM and democratise access to data (17. Kolloquium Investor - Hochschule -

- Bauindustrie)," 2024. [Online]. Dostupné na: <https://www.bim.bayern.de/wp-content/uploads/2023/06/Kolloquium-17-TUM-Baprozessmanagment-und-Bau-Bauindustrie.pdf>.
- [122] Д. Hill, D. Foldesi, S. Ferrer, M. Friedman, E. Loh a F. Plaschke, "Solving the construction industry productivity puzzle", 2015. [Online]. Dostupné na: <https://www.bcg.com/publications/2015/engineered-products-project-business-solving-construction-industrys-productivity-puzzle>.
- [123] "SCOPE - Projektdatenumgebung und Modellierung multifunktionaler Bauprodukte mit Fokus auf die Gebäudehülle," 1. ledna 2018. [Online]. Dostupné na: <https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html>. [Datum přístupu: 2. března 2025].
- [124] Apple.com, "Pixar, Adobe, Apple a NVIDIA vytvořily Alianci pro OpenUSD, která bude prosazovat otevřené standardy pro 3D obsah," 1. srpna 2023. [Online]. Dostupné na: <https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/>. [Přístup: 2. března 2025].
- [125] AECmag, "ADSK's granular data strategy", 25. července 2024. [Online]. Dostupné na: <https://aecmag.com/technology/autodesks-granular-data-strategy/>. [Datum přístupu: 15. března 2025].
- [126] A. Bojko, "Doba změn: IFC je minulostí aneb proč se ADSK a další dodavatelé CAD chtějí vzdát IFC kvůli USD ve 14 klíčových faktech", 24. 11. 2024. [Online]. Dostupné na: <https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82cccd10a>. [Datum oslovení: 23 fevryall 2025].
- [127] A. Boyko, "ENG BIM Cluster 2024 | Bitva o data a aplikace LLM a ChatGPT ve stavebnictví", 7. srpna 2024. [Online]. Dostupné: ENG BIM Cluster 2024 | Bitva o data a uplatnění LLM a ChatGPT ve stavebnictví. [Dostupné: 15. března 2025].
- [128] "Jeffrey Zeldman představuje", 6. května 2008. [Online]. Dostupné na: <https://zeldman.com/2008/05/06/content-precedes-design/>. [Datum přístupu: 15. března 2025].
- [129] A. Boyko, "DWG Analyse with ChatGPT | DataDrivenConstruction", 5. března 2024. [Online]. Dostupné na: <https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datadrivenconstruction>. [Datum odkazu: 15. března 2025].
- [130] McKinsey , "Průvodce McKinsey pro překonání konkurence ve věku digitální a umělé inteligence", 2023. [Online]. Dostupné na: <https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired>. [Datum oslovení: 30. března 2025].

- [131] Forbes, "Data Storytelling: The Essential Data Science Skill Everyone Needs", 31. března 2016. [Online]. Dostupné na: <https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/>. [Datum přístupu: 15. března 2025].
- [132] J. Bertin, "Grafika a grafické zpracování informací", 8. září 2011. [Online]. Dostupné na: https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=csqX_xnm4tcC&redir_esc=y. [Datum přístupu: 15. března 2025].
- [133] CauseWeb, "Wells/Wilks on Statistical Thinking," [online]. Dostupné na: <https://www.causeweb.org/cause/resources/library/r1266>. [Datum přístupu: 15. března 2025].
- [134] Ministrymagazine, "Jak věda objevila stvoření", leden 1986. [Online]. Dostupné na: <https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation>. [Datum odkazu: 15. března 2025].
- [135] BCG, "Data-Driven Transformation: Accelerate at Scale Now", 23. května 2017. [Online]. Dostupné na: <https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation>. [Přístup 15. května 2024].
- [136] "Jak vytvořit datovou architekturu, která bude hnacím motorem inovací - dnes i zítra", 3. června 2020. [Online]. Dostupné na: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow>. [Datum vystoupení: 15. března 2025].
- [137] Oxford, "Woodrow Wilson 1856-1924," [Online]. Available: <https://www.oxfordreference.com/display/10.1093/acref/9780191866692.001.0001/q-oro-ed6-00011630>. [Datum vystoupení: 15. března 2025].
- [138] "Convertors", 2024. [Online]. Dostupné na: <https://datadrivenconstruction.io/index.php/convertors/>.
- [139] PWC, "Sizing the prize What's the real value of AI for your business and how can you capitalise?", 1. ledna 2017. [Online]. Dostupné na: <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>. [Datum přístupu: 18. února 2025].
- [140] "Potrubí ve výstavbě", 2024. [Online]. Dostupné na: <https://datadrivenconstruction.io/index.php/pipeline-in-construction/>.
- [141] Wikipedia, "Apache NiFi", 1. ledna 2025. [Online]. Dostupné na: https://de.wikipedia.org/wiki/Apache_NiFi. [Datum přístupu: 5. března 2025].
- [142] n8n, "Gmail AI Auto-Responder: Create Draft Replies to incoming emails", 1. května 2024. [Online]. Dostupné na: <https://n8n.io/workflows/2271-gmail-ai-auto-responder-create-draft>.

- replies-to-incoming-emails/. [Přístup 15. března 2025].
- [143] n8n, "Real Estate Daily Deals Automation with Zillow API, Google Sheets and Gmail", 1. března 2025. [Online]. Dostupné na: <https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with-zillow-api-google-sheets-and-gmail/>. [Datum přístupu: 15. března 2025].
- [144] B. T. O'Neill, "Míra neúspěšnosti projektů v oblasti analytiky, umělé inteligence a velkých dat = 85 % - jáje!", 1. ledna 2025. [Online]. Dostupné na: <https://designingforanalytics.com/resources/failure-rates-for-analytics-bi-iot-and-big-data-projects-85-yikes/>.
- [145] J. Neyman, On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection, Oxford University Press, 1934.
- [146] T. J. S. a J. S. Jesse Perla, "A Problem that Stumped Milton Friedman", Quantitative Economics with Julia, 1. 1. 2025. [Online]. Dostupné na: https://julia.quantecon.org/dynamic_programming/wald_friedman.html. [Datum přístupu: 1. května 2024].
- [147] T. Landsall-Welfair, Předpověď aktuální nálady národa, Significance, 2012.
- [148] A. Bojko, "San Francisco. Stavebnictví 1980-2019," 2024. [Online]. Dostupné na: <https://www.kaggle.com/search?q=San+Francisco.+Building+sector+1980-2019>.
- [149] A. Boyko, "Kaggle: RVT IFC Files 5000 Projects", 2024. [Online]. Dostupné na: <https://www.kaggle.com/datasets/artemboiko/rvtifc-projects>.
- [150] CFMA, "Příprava na budoucnost s propojeným stavebnictvím". [Online]. Dostupné na: <https://cfma.org/articles/preparing-for-the-future-with-connected-construction>. [Přístup 15. března 2025].
- [151] Cisco, "Průzkum společnosti Cisco odhalil, že téměř tři čtvrtiny projektů internetu věcí selhávají", 22. května 2017. [Online]. Dostupné na: <https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2017/m05/cisco-survey-reveals-close-to-three-fourths-of-iot-projects-are-failing.html>.
- [152] "Podmínky nutné pro uchování rostlinných fosilií", 2024. [Online]. Dostupné na: <https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html>.
- [153] "Fink z BlackRock o dluhopisech, fúzích a akvizicích, recessi v USA a volbách: celý rozhovor," 2023. [Online]. Dostupné na: <https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video>.
- [154] cio, "12 slavných katastrof AI", 02. října 2024. [Online]. Dostupné na:

- <https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html>. [Datum přístupu: 15. března 2025].
- [155] G. Kasparov, Deep Thinking, PublicAffairs, 2017.
- [156] Wikipedia, "Kaggle", 1. ledna 2025. [Online]. Dostupné na: <https://en.wikipedia.org/wiki/Kaggle>. [Datum přístupu: 15. března 2025].
- [157] Kaggle, "Titanic - strojové učení z katastrofy", 1. ledna 2025. [Online]. Dostupné na: <https://www.kaggle.com/competitions/titanic/overview>. [Dostupné 10. března 2025].
- [158] Ш. Johri, "Creating ChatGPT: From Data to Dialogue", 2024. [Online]. Dostupné na: <https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/>.
- [159] П. Domingos, "A few useful things to know about machine learning", 2024. [Online]. Dostupné na: <https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf>.
- [160] J. Saramago, "Citát," [online]. Dostupné na: <https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered>. [Datum přístupu: 17. března 2025].
- [161] NVIDIA, "Enhance Your Training Data with New NVIDIA NeMo Curator Classifier Models", 19. prosince 2024. [Online]. Dostupné na: <https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/>. [Přístup 25. března 2025].
- [162] "Společnost NVIDIA oznámila zásadní vydání modelů nadace Cosmos World Foundation a nástrojů pro fyzikální umělou inteligenci," 18. března 2025. [Online]. Dostupné na: <https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools>. [Datum přístupu: 25. března 2025].
- [163] NVIDIA, "NVIDIA Isaac Sim," [Na internetu]. Dostupné na: <https://developer.nvidia.com/isaac/sim>. [Přístup 25. března 2025].
- [164] M. Quarterly, "Proč digitální strategie selhávají", 25. ledna 2018. [Online]. Dostupné na: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail>. [Datum přístupu: 15. března 2025].
- [165] M. J. Perry, "Moje oblíbené citáty Miltona Friedmana", 17. listopadu 2006. [Online]. Dostupné na: <https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/>. [Datum přístupu: 1. března 2025].
- [166] J. A. Wheeler, "Informace, fyzika, kvantová fyzika: hledání souvislostí", 1990.
- [169] A. Boyko, "Lobbying wars and BIM development. Část 5: BlackRock je páñem všech technologií. Jak korporace ovládají otevřený zdrojový kód," 2024. [Online]. Dostupné na:

- <https://boikoartem.medium.com/lobbyist-wars-and-the-development-of-bim-d72ad0111a7d>.
- [170] T. Krijnen a J. Beetz, "A SPARQL query engine for binary-formatted IFC building models", *Advanced Engineering Informatics*, 2024.
- [171] "Počet podniků v odvětví stavebnictví ve Spojeném království v roce 2021 podle velikosti podniku", 2024. [Online]. Dostupné na: <https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/>.
- [172] "5 000 projektů IFC&RVT," 2024. [Online]. Dostupné na: <https://www.kaggle.com/code/artemboiko/5000-projects-ifc-rvt-datadrivenconstruction-io>.
- [173] M. Popova, "It from Bit: Pioneering Physicist John Archibald Wheeler on Information, the Nature of Reality, and Why We Live in a Participatory Universe", 2008. [Online]. Dostupné na: <https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/>. [Datum odkazu: únor 2025].
- [174] *Lobbistické války o data ve stavebnictví / Technofeudalismus a historie skryté minulosti BIM*. [Film]. Německo: Artyom Boiko, 2023.
- [175] A. Boyko, "CHATGPT WITH REVIT AND IFC | Automatické vyhledávání dokumentů a dat z projektů", 16. listopadu 2023. [Online]. Dostupné na: https://www.youtube.com/watch?v=ASXolti_YPs&t. [Datum přístupu: 2. března 2025].
- [176] M. & Company, "Three new mandates for capturing a digital transformation's full value", 22. 1. 2022. [Online]. Dostupné na: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value>. [Datum přístupu: 15. února 2025].
- [177] KPMG, "Stavebnictví v digitálním světě", 1. května 2021. [Online]. Dostupné na: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-cs.pdf>. [Datum přístupu: 5. dubna 2024].
- [178] LLP, KPMG, "Cue Construction 4.0: Make-or-Break Time", 17. března 2023. [Online]. Dostupné na: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [Datum přístupu: 15. února 2025].
- [179] O. Business, "Satya Nadella prozradil, jak AI agenti naruší modely SaaS", 10. ledna 2025. [Online]. Dostupné na: <https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models>. [Datum přístupu: 15. března 2025].
- [180] Forbes, "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says", 23. března 2016. [Online]. Dostupné na: <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming>

- least-enjoyable-data-science-task-survey-says/. [Datum přístupu: 15. března 2025].
- [181] Ministerstvo zahraničních věcí, Commonwealthu a rozvoje Spojeného království, "Strategie digitálního rozvoje 2024 až 2030", 18. března 2024. [Online]. Dostupné na: <https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030>. [Datum přístupu: 15. března 2025].
- [182] "Vize a strategie v oboru navrhování budov," 7. listopadu 2003. [Online]. Dostupné na: <https://web.archive.org/web/20030711125527/http://usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112>. [Datum vystoupení: 5. března 2025].
- [183] M. Bocharov, "Informační modelování", březen 2025. [Online]. Dostupné na: <https://www.litres.ru/book/mihail-evgenievich-bocharov/informacionnoe-modelirovaniye-v-rossii-71780080/chitat-onlayn/?page=5>. [Datum vystoupení: 15. března 2025].
- [184] "Integrated Design-Through-Manufacturing: Benefits and Rationale", 2000. [Online]. Dostupné na: https://web.archive.org/web/20010615093351/http://www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf. [Datum vydání: 25. března 2025].
- [185] CAD Vendor, "Program Open BIM je marketingová kampaň," 12. března 2012. [Online]. Dostupné na: <https://web.archive.org/web/20120827193840/http://www.graphisoft.com/openbim/>. [Datum oslovení: 30. března 2025].

INDEX PŘEDMĚTŮ

3D, 8, 14, 71, 73, 84, 191, 210, 215, 232, 234, 263, 276, 277, 287, 298, 299, 302, 303, 306, 307, 337, 375, 393, 394, 448, 468, 480

4D, 84, 172, 196, 199, 210, 229, 234, 237, 287

4IR, 43

5D, 84, 172, 196, 210, 237, 287, 492

6D, 172, 196, 229, 232, 233, 234, 235

7D, 232, 233, 234, 287

8D, 172, 196, 229, 232, 233, 234, 287

A

AI, 3, 50, 52, 100, 102, 103, 106, 107, 116, 395, 457, 459, 461, 477

AIA, 289

AIM, 3, 289

AMS, 14, 84, 151, 153

Apache Airflow, 311, 361, 362, 363, 364, 366, 367, 369, 371, 399, 494

Apache NiFi, 116, 311, 361, 367, 368, 371, 399, 482, 494

Apache ORC, 62, 330, 378

Apache Parquet, 62, 67, 380, 381, 494

API, 54, 94, 95, 97, 109, 137, 138, 154, 168, 218, 219, 256, 257, 260, 271, 273, 294, 295, 296, 297, 300, 308, 326, 329, 342, 366, 369, 370, 488

B

BDS, 258, 259

Big Data, 9, 67, 245, 383

BIM, 2, 3, 4, 6, 3, 14, 17, 24, 56, 58, 60, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 84, 92, 137, 138, 139, 140, 141, 142, 144, 150, 154, 156, 166, 172, 183, 184, 186, 187, 190, 195, 196, 208, 210, 211, 213, 216, 217, 218, 219, 220, 221, 222, 227, 228, 237, 238, 239, 242, 243, 246, 250, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 266, 271, 272, 273, 275, 276, 277, 278, 279, 280, 282, 285, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 308, 309, 324, 328, 337, 351, 356, 358, 361, 367, 372, 376, 383, 389, 402, 403, 405, 413, 416, 431, 448, 462, 466, 475, 476, 480, 492, 494, 497

475, 476, 492, 494, 497

BlackBox, 240, 242, 243.

BMS, 8

Bokeh, 320, 337

BOM, 76, 77, 79, 257, 263

Vymezovací pole, 234, 373, 392, 393, 394, 414, 415, 480

BREP, 142, 234, 263, 264, 276, 283, 284

C

6, 14, 18, 24, 56, 57, 58, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 84, 85, 95, 95, 97, 111, 126, 137, 138, 139, 140, 141, 142, 144, 146, 147, 152, 153, 155, 156, 166, 172, 175, 183, 184, 186, 187, 190, 195, 196, 206, 208, 210, 211, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 224, 227, 228, 232, 234, 237, 238, 239, 242, 243, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 303, 308, 309, 324, 328, 337, 344, 351, 356, 358, 361, 367, 372, 376, 383, 389, 402, 403, 405, 413, 416, 431, 448, 462, 466, 475, 476, 480, 492, 494, 497

CAE, 16, 283

CAFM, 14, 24, 62, 84, 151, 153, 172, 233, 278, 326, 356, 387, 458, 473

CAM, 16, 78

CAPEX, 14, 82

CDE, 84, 175, 388, 389, 390

ChatGPT, 103, 104, 109, 110, 124, 129, 221, 303, 307, 343, 349, 421

Claude, 103, 104, 109, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443

CO₂, 72

CO₂, 229, 234, 235, 236, 237, 238, 239, 243

COBie, 156, 289, 292.

CoE, 56, 168, 169, 170, 477

Kopilot, 114, 116, 459

CPIXML, 143, 272, 273, 276, 277, 279, 285, 296

CPM, 14, 17, 62, 166, 175, 233, 326, 473

CQMS, 14, 84, 177, 178, 462

CRM, 109, 369, 459.

CRUD, 51, 459

CSG, 263

CSV, 61, 62, 88, 89, 120, 123, 128, 129, 130, 131, 135, 168,

268, 272, 280, 329, 333, 344, 354, 354, 356, 373, 377,

378, 379, 380, 401, 407, 414, 474, 480

D

DAE, 276, 277, 278, 280, 281, 284, 285, 414
 DAG, 362, 363, 365, 366
 Dash, 320, 336, 337.
 Správa dat, 373, 395, 396, 398, 400, 401
 Data Lake, 214, 373, 376, 384, 385, 386, 387, 388, 389, 390, 400
 Data Lakehouse, 67, 373, 386, 387, 388
 Datový minimalismus, 373, 395, 396, 400, 401
 Data Swamp, 373, 395, 397, 401
 Datový sklad, 382, 383, 400
 Data jako služba, 487
 na základě dat, 50, 170, 460, 461, 484, 486
 DataFrame, 67, 117, 121, 122, 123, 125, 129, 130, 131, 133, 134, 135, 137, 219, 220, 224, 237, 308, 328, 329, 330, 332, 333, 344, 345, 347, 348, 354, 365, 377, 381, 409, 414, 415, 475
 DataOps, 170, 373, 398, 399, 400, 401, 475
 DeepSeek, 103, 104, 107, 109, 110, 124, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494
 DGN, 8, 140, 186, 227, 357
 DWG, 8, 70, 71, 73, 97, 140, 186, 211, 227, 272, 287, 302, 303, 304, 307, 357, 376, 497
 DWH, 67, 373, 376, 382, 383, 384, 386, 387, 388, 389, 390
 DXF, 8, 73, 277

E

ECM, 58, 175
 ECS, 142
 EIR, 289
 eLOD, 289
 ELT, 384, 385
 EPM, 14, 166, 198
 ERP, 2, 11, 12, 14, 17, 18, 24, 25, 58, 62, 109, 153, 166, 172, 175, 196, 198, 210, 232, 239, 240, 241, 242, 243, 244, 245, 246, 247, 249, 272, 277, 278, 279, 282, 326, 351, 356, 361, 369, 387, 388, 389, 390, 458, 462, 473, 475, 487, 492
 ESG, 196, 235, 236, 238
 ETL, 6, 1.1-8, 32, 49, 81, 113, 116, 119, 128, 188, 193, 219, 291, 311, 312, 317, 323, 324, 325, 326, 327, 328, 330, 331, 333, 338, 339, 340, 343, 344, 348, 349, 350, 351, 353, 354, 356, 361, 362, 363, 364, 365, 367, 371, 372, 381, 382, 383, 384, 385, 399, 409, 430, 475, 481, 494
 Excel, 57, 61, 62, 65, 66, 85, 88, 111, 120, 123, 123, 125, 132, 154, 167, 187, 210, 223, 224, 226, 227, 228, 278, 291, 329, 333, 342, 351, 365, 376, 378, 459, 475, 497.
 Výpis, 81, 128, 134, 193, 311, 323, 324, 326, 328, 330, 345, 361, 362, 363, 365, 383, 384, 450, 475

F

Perí, 62, 123, 330, 378
 FPDF, 339, 340, 341, 342, 343

G

GDPR, 109
 GIS, 58
 GLTF, 143, 278
 Tabulky Google, 368, 370
 Grok, 103, 104, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443

H

HDF5, 62, 67, 123, 329, 330, 378, 379, 380
 HiPPO, 29, 37, 95, 424, 477, 484, 490
 HTML, 123, 340, 365, 370

I

IDS, 289, 290, 291.
 IFC, 8, 73, 138, 142, 186, 227, 261, 262, 263, 264, 265, 266, 267, 268, 272, 273, 273, 276, 277, 278, 279, 280, 284, 286, 292, 296, 302, 329, 357, 414, 415, 417, 417, 497
 IGES, 262, 263, 276.
 iLOD, 289
 IoT, 10, 18, 67, 271, 367, 369, 405, 413, 417, 417, 418, 419, 455, 460, 465, 482, 484.
 ISO 19650, 388

J

JavaScript, 320, 369, 378
 JSON, 88, 89, 90, 92, 123, 128, 142, 269, 272, 280, 329, 330, 333, 378, 474, 480
 Jupyter Notebook, 114, 115, 116, 130, 187, 224, 330, 346, 417, 425

K

Kaggle, 115, 121, 130, 187, 224, 303, 307, 330, 346, 408, 415, 417, 425, 426, 430, 431, 433
 k-NN, 392, 393, 442, 445, 446, 447, 448
 KPI, 245, 311, 317, 318, 319, 320, 321, 324, 353, 372, 389, 478, 479

L

LEED, 235, 236, 238.
 LlaMa, 103, 104, 120, 124, 129, 133, 166, 219, 300, 329,

343, 349, 381, 421, 429, 443, 494
 LLM, 3, 4, 24, 29, 50, 51, 52, 55, 56, 92, 95, 99, 102, 103,
 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114,
 115, 116, 118, 120, 124, 125, 126, 129, 131, 133, 134,
 135, 166, 187, 219, 220, 221, 222, 222, 224,
 225, 226, 231, 237, 238, 251, 294, 300, 301, 302, 303,
 304, 305, 306, 307, 308, 309, 328, 329, 332, 333, 334,
 335, 336, 338, 339, 340, 343, 344, 345, 346, 347, 348,
 349, 354, 370, 372, 376, 381, 382, 392, 394, 399, 413,
 425, 429, 430, 432, 433, 435, 443, 444, 457, 460, 461,
 476, 488, 494
 Zatížení, 81, 128, 134, 193, 311, 323, 325, 326, 333, 334,
 338, 339, 342, 343, 361, 362, 363, 365, 367, 383, 384,
 450, 475
 LOD, 287, 289
 LOI, 287
 LOMD, 287
 Low-Code, 368, 369

M

Matplotlib, 123, 306, 320, 333, 335, 342, 372, 409, 411,
 494
 MCAD, 77, 78, 257, 284
 MEP, 14, 175, 492
 SÍŤOVINA, 234, 283, 284, 285, 296, 462
 Microsoft SQL, 65
 Mistral, 103, 104, 107, 110, 129, 133, 166, 219, 221, 300,
 329, 343, 349, 381, 421, 429, 443, 494
 MRP, 11, 12
 MS Project, 70
 MySQL, 63, 64, 65, 291, 329, 330

N

n8n, 116, 311, 361, 368, 369, 370, 371, 372
 NLP, 69, 394
 No-Code, 368, 369
 NURBS, 142, 282, 283, 284, 285
 NWC, 8, 276

O

OBJ, 143, 273, 276, 277, 278, 280, 281, 284, 285
 OCCT, 273
 OCR, 69, 128, 132, 134
 OmniClass, 154, 155, 156.
 otevřený BIM, 142, 216, 219, 256, 261, 278, 291
 Open Source, 45, 55, 97, 98, 107, 108, 273, 275, 494
 SOVA, 267, 268, 269.

P

Pandy, 56, 67, 103, 117, 118, 119, 120, 121, 122, 123, 125,
 130, 134, 138, 186, 187, 220, 221, 225, 227, 269, 300,
 303, 328, 329, 330, 377, 378, 380, 381, 403, 408, 409,
 410, 411, 412, 414, 475, 479, 494
 Parkety, 67, 123, 329, 330, 373, 378, 379, 380, 381, 382,
 386, 401, 403, 414, 474, 480
 PDF, 69, 70, 71, 85, 111, 126, 127, 128, 129, 130, 131, 132,
 134, 146, 147, 177, 186, 190, 191, 211, 215, 278, 338,
 339, 340, 341, 342, 342, 344, 345, 346, 347, 354, 355,
 356, 357, 365, 376, 475, 480, 497
 PDM, 16
 PHP, 63, 339
 Potrubí, 44, 53, 115, 128, 183, 303, 307, 308, 311, 312,
 349, 350, 351, 352, 354, 355, 356, 367, 370, 417, 430,
 476
 PLM, 16, 246
 PLN, 8, 272, 296
 Plotly, 320, 336, 337, 372
 PMIS, 3, 24, 32, 62, 152, 196, 198, 211, 239, 240, 245, 246,
 247, 248, 249, 250, 272, 326, 351, 387, 388, 389, 390,
 458, 462, 487
 PMS, 84, 151, 233
 PostgreSQL, 63, 64, 65, 329, 395
 Power BI, 320, 372, 482
 soukromý kapitál, 464, 487
 Python, 56, 63, 103, 105, 112, 114, 115, 117, 118, 119, 129,
 130, 131, 133, 134, 135, 166, 167, 179, 187, 219, 224,
 225, 227, 303, 304, 320, 329, 330, 332, 333, 339,
 340, 346, 348, 356, 367, 369, 372, 378, 409, 417, 426,
 459, 475, 476, 479, 494.

Q

QTO, 72, 196, 214, 215, 216, 217, 218, 219, 221, 223, 225,
 226, 228, 237, 238, 242, 243, 251, 301, 302, 475, 479
 QWEN, 103, 104, 124, 129, 133, 166, 219, 221, 300, 329,
 349, 343, 349, 381, 429, 443

R

RAG, 111, 116
 RDBMS, 63, 64, 65, 82, 89.
 RDF, 267, 268, 269
 RegEx, 126, 136, 177, 179, 331, 332, 333, 357, 474
 RFID, 8, 18, 58, 84, 418, 419, 455, 460, 465, 482
 ROI, 311, 317, 319, 321, 351, 370, 389, 479
 RPM, 14, 84, 151, 331, 333, 334
 RVT, 8, 73, 77, 140, 186, 227, 272, 296, 300, 302, 322, 329,
 357, 376, 414, 415, 417, 497

S

SaaS, 24, 50, 51, 52, 458
ROZSAH, 273, 277
SDK, 139, 141, 257, 264, 273, 281, 286, 296, 329, 414
Seaborn, 123, 320, 336, 337, 408, 412
SPARQL, 269
SQL, 63, 65, 66, 88, 89, 103, 105, 119, 123, 166, 168, 268, 269, 279, 276, 277, 300, 329, 376, 392, 494
SQLite, 63, 64, 65, 166, 167, 296, 329, 363
KROK, 261, 262, 263, 266, 272, 276, 277, 292
Streamlit, 336, 337
SVF, 142, 276, 285

T

Transformace, 128, 193, 311, 323, 325, 326, 330, 331, 333, 334, 340, 341, 342, 344, 347, 361, 362, 363, 365, 367, 383, 384, 450, 475

U

Uniclass, 154, 155, 156.
USD, 142, 143, 255, 276, 277, 278, 279, 280, 281, 284, 285, 286, 292, 296

V

VectorOps, 373, 398, 400, 401
VR, 84, 271, 285

W

WhiteBox, 240, 242, 243.

X

XLSX, 8, 61, 62, 123, 128, 129, 231, 268, 276, 277, 280, 296, 299, 308, 351, 373, 377, 378, 379, 380, 401, 474, 480
XML, 61, 88, 89, 92, 128, 143, 269, 272, 277, 280, 280, 291, 299, 329, 378, 414, 474